

### Protein Sequence Searches - 10/8/04

All of the sequence databases on Compugen/ABSS have been updated. A change has occurred in the protein databases.

- Two protein databases, SPTREMBL and SwissProt, are now produced as a single, merged database called UniProt.
- Results from UniProt have the file extension **.rup**.
- Sequences in UniProt are identified by the same ID that had been used in SPTREMBL or SwissProt.

If you have any questions regarding these changes or your results, please contact any STIC searcher.

Blank

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 120.835 Seconds  
(without alignments)  
5233.063 Million cell updates/sec

Title: SEQ1-A  
Perfect score: 5835  
Sequence: 1 MAWKILPIYILLLLSVFIQ.....DMDYLPVFNQGIINPMLS 1099

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.1

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Uniprot 02:.\*  
1: uniprot\_sprot:.\*  
2: uniprot\_trembl:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	5820.9	99.8	1404	2	Q92954	Q92954 homo sapien
2	5797.9	99.4	1404	2	Q9BX49	Q9BX49 homo sapien
3	3417	58.6	933	2	Q6ZMZ5	Q6ZMZ5 homo sapien
4	3417	58.6	933	2	BAD18580	BAD18580 h cdna fl
5	2622	44.9	1054	2	Q9JMS9	Q9JMS9 mus musculu
6	1355.4	23.2	1225	2	Q9VRA9	Q9VRA9 drosophila
7	1354.4	23.2	5173	1	MUC2_HUMAN	Q02817 homo sapien
8	1337.7	22.9	1761	2	Q7KTF6	Q7KTF6 drosophila
9	1337.7	22.9	1761	2	AAS64673	AAS64673 drosophila
10	1286	22.0	3150	2	Q7PMD5	Q7PMD5 anopheles g
11	1185.8	20.3	3409	2	Q6SSB6	Q6SSB6 chlamydomon
12	1185.8	20.3	3409	2	AAS07044	AAS07044 chlamydomon
13	1143.3	19.6	1664	1	SLP1_CLOTH	Q06852 clostidium
14	1117.1	19.1	1349	2	Q9WQ4	Q9WQ4 homo sapien
15	1112.3	19.1	3432	2	Q9IR51	Q9IR51 drosophila
16	1112.3	19.1	3458	2	Q9IR52	Q9IR52 drosophila
17	1110.2	19.0	3889	2	Q6SSS8	Q6SSS8 chlamydomon
18	1110.2	19.0	3889	2	AAS07042	AAS07042 chlamydomon
19	1087.6	18.6	1795	2	Q76894	Q76894 drosophila
20	1072	18.4	9234	2	Q7KTF5	Q7KTF5 drosophila
21	1072	18.4	9234	2	AAN10531	AAN10531 drosophila
22	1063.3	18.2	1079	2	Q9N4S7	Q9N4S7 caenorhabdi
23	1034.5	17.7	2284	2	Q9VPG1	Q9VPG1 drosophila
24	1027.6	17.6	23015	2	Q8IQ18	Q8IQ18 drosophila
25	1027.6	17.6	23015	2	AAN10358	AAN10358 drosophila
26	1024	17.5	1607	2	Q8H6QS	Q8H6QS phytoptior
27	1024	17.5	1607	2	AAP74661	AAP74661 phytoptior
28	1020.5	17.5	972	2	Q7QKK7	Q7QKK7 anopheles g
29	1015.5	17.4	1274	2	Q2Q007	Q2Q007 caenorhabdi
30	1014.9	17.4	2187	2	P70670	P70670 mus musculu
31	1014.7	17.4	10578	2	Q8ISF5	Q8ISF5 caenorhabdi

32 1014.7 17.4 18519 2 Q8ISF6  
33 1014.7 17.4 18534 2 Q8ISF7  
34 1013.8 17.4 34350 2 Q8WZ42  
35 1009.9 17.3 1489 2 Q96449  
36 1008.4 17.3 7962 2 Q10465  
37 998.1 17.1 5703 1 MUSB\_HUMAN  
38 996.1 17.1 2112 2 Q9VEL9  
39 991.2 17.0 1480 2 Q9LIE8  
40 982.1 16.8 2042 2 Q767L8  
41 982.1 16.8 2042 2 BAD08434  
42 967.7 16.6 926 2 Q9VVG2  
43 966.6 16.6 1121 2 Q7Z884  
44 965.6 16.5 3166 2 Q9W3Z0  
45 963.4 16.5 1720 2 Q8I486

## ALIGNMENTS

RESULT 1  
Q92954 PRELIMINARY; PRT; 1404 AA.  
ID AC Q92954;  
DT 01-FEB-1997 (TREMELrel. 02, Created)  
DT 01-FEB-1997 (TREMELrel. 02, Last sequence update)  
DT 01-MAR-2004 (TREMELrel. 26, Last annotation update)  
DE Megakaryocyte stimulating factor.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,  
RA Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,  
RA Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,  
RA Bhatia S., Kriz R., Hewick R., Clark S.C.;  
RT "Purification, Biochemical Characterization, and Cloning of a Novel  
RT Megakaryocyte Stimulating Factor that has Megakaryocyte Colony  
RT Stimulating Activity.";  
RL Blood 78:279-279(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Merberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P.,  
RA Fitzgerald M., Scaltreto J., Kelleher K., Preissner K., Kriz R.,  
RA Jacobs K., Turner K.;  
RT "A Comparison of Vitronectin and Megakaryocyte Stimulating Factor.";  
RL (In) Preissner K.T., Rosenblatt S., Kost C., Wegerhoff J.,  
RL Mosher D.F. (eds.);  
RL BIOLOGY OF VITRONECTINS AND THEIR RECEPTORS., pp.45-52, Elsevier  
RL Science Publishers B.V. (1993).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,  
RA Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,  
RA Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,  
RA Bhatia S., Kriz R., Hewick R., Clark S.C.;  
RT Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
RL EMBL; U70136; AAB09089.1; -.  
DR HSSP; P04004; 10C0.  
DR Genew; HGNC:9364; PRG4.  
DR GO; GO:0008283; P:cell proliferation; TAS.  
DR InterPro; IPR000585; Hemopexin.  
DR InterPro; IPR001212; Somatomedin\_B.  
DR Pfam; PF00045; Hemopexin; 2.  
DR Pfam; PF01033; Somatomedin B; 2.  
DR PRINTS; PR00022; SOMATOMEDINB.  
DR SMART; SM00120; HX; 2.  
DR SMART; SM00201; SO; 2.  
DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN 1.  
DR PROSITE; PS00524; SOMATOMEDIN B; 2.  
SQ SEQUENCE 1404 AA; 151090 MW; AABD7AD19B35F4F6 CRC64;

```
Query Match      99.8%; Score 5820.9; DB 2; Length 1404;
Best Local Similarity 96.4%; Pred. No. 4.2e-103;
Matches 1099; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQOVSSQ----- 25
DB 1 MAWKTLPIYLLLLSVFVIQOVSSQDLSSCAGRCGEGYSRDATCNCYHMECCPDF 60
QY 26 -----ELSCKGRCFESFERGECDDAOCKKYDKCCPDYEFSCAEVHNPTSPSSKKAP 79
DB 61 KRVCCTAELSCKGRCFESFERGECDDAOCKKYDKCCPDYEFSCAEVHNPTSPSSKKAP 120
QY 80 PPSGASQTIKSTTKRSPKPPNKKTKVIESEITEHSVSENQESSSSSSSSSSSTIW 139
DB 121 PPSGASQTIKSTTKRSPKPPNKKTKVIESEITEHSVSENQESSSSSSSSSSSTIW 180
QY 140 KIKSSKNSAANRELQKLVKONKNRTKKPTPKPPVVDAGSLDNGDFKVTTPDTST 199
DB 181 KIKSSKNSAANRELQKLVKONKNRTKKPTPKPPVVDAGSLDNGDFKVTTPDTST 240
QY 200 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSILVNKETTIVETKETTINKQSTSDG 259
DB 241 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSILVNKETTIVETKETTINKQSTSDG 300
QY 260 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTTPKEPAS 319
DB 301 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTTPKEPAS 360
QY 320 TTPKEPTTTPKSAPTTPKEPATTTKSAPTTPKEPATTTKBPATTTKBPATTTKBP 379
DB 361 TTPKEPTTTPKSAPTTPKEPATTTKSAPTTPKEPATTTKBPATTTKBPATTTKBP 420
QY 380 APTTTSAPTTTPKEPATTTTPKBPATTTTPKBPATTTTPKEPATTTTPKEPATTPK 439
DB 421 APTTTSAPTTTPKEPATTTTPKBPATTTTPKBPATTTTPKEPATTTTPKEPATTPK 480
QY 440 EPAPTAPKBPATTTTPKEPATTTTPKBPATTTTPKBPATTTTPKBPATTTTPKBPAT 499
DB 481 EPAPTAPKBPATTTTPKEPATTTTPKBPATTTTPKBPATTTTPKBPATTTTPKBPAT 540
QY 500 TTKSAPTTTPKBPSTTTTPKEPATTTTPKBPATTTTPKBPATTTTPKBPATTTTPKBP 559
DB 541 TTKSAPTTTPKBPSTTTTPKEPATTTTPKBPATTTTPKBPATTTTPKBPATTTTPKBP 600
QY 560 APTAPKEPATTTTPKEPATTTTPKBPATTTTPKBPATTTTPKBPATTTTPKBPATTT 619
DB 601 APTAPKEPATTTTPKEPATTTTPKBPATTTTPKBPATTTTPKBPATTTTPKBPATTT 660
QY 620 PEPAPTTTPKAAAPNTPKBPATTTTPKBPATTTTPKBPATTTTPKBPATTTTPKBPAT 679
DB 661 PEPAPTTTPKAAAPNTPKBPATTTTPKBPATTTTPKBPATTTTPKBPATTTTPKBPAT 720
QY 680 APTTPKBPAPKELAPTTTKEPTSTSDKBPATTTKGTAPTTTPKEPATTTTPKBPATTPKG 739
DB 721 APTTPKBPAPKELAPTTTKEPTSTSDKBPATTTKGTAPTTTPKEPATTTTPKBPATTPKG 780
QY 740 TAPTTLKEPATTTTPKBPAPKELAPTTTKEPTSTSDKBPATTTKGTAPTTTPKEPATTTTPK 799
DB 781 TAPTTLKEPATTTTPKBPAPKELAPTTTKEPTSTSDKBPATTTKGTAPTTTPKEPATTTTPK 840
QY 800 KPAPTTPPEPTTTPSVSTPTTTTKEPTTTHKSPDESTPELSABPTPKALENSKPEPCVPT 859
DB 841 KPAPTTPPEPTTTPSVSTPTTTTKEPTTTHKSPDESTPELSABPTPKALENSKPEPCVPT 900
QY 860 TKTPAATKPEMTTAKDKTTERDLRTTPETTTAAAPKMTKETATTTKTESKITATTQV 919
DB 901 TKTPAATKPEMTTAKDKTTERDLRTTPETTTAAAPKMTKETATTTKTESKITATTQV 960
QY 920 TSTTTQDTPFKITLLKLTLLAPKVTITTKTITTTTTEIMNKPBEATAKPKDRATNSKATTPK 979
DB 961 TSTTTQDTPFKITLLKLTLLAPKVTITTKTITTTTTEIMNKPBEATAKPKDRATNSKATTPK 1020
QY 980 PQKPTKAPKPTSTKPKTWPVRVKPTTTPRKMSTMTPELNPTSRIAEAMLQTTTRPN 1039
```

```
Db 1021 FOKPTKAPKPTSTTKPKTMRVRKPTTTPRKMSTMTPELNPTSRIAEAMLQTTTRN 1080
QY 1040 QTPNSKLVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMDYLPVNPQGIINPMLS 1099
DB 1081 QTPNSKLVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMDYLPVNPQGIINPMLS 1140

RESULT 2
Q9BX49 PRELIMINARY; PRT; 1404 AA.
AC Q9BX49;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE BG17416.2 (MSF: megakaryocyte stimulating factor ).
GN Name=BG17416.2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wray P.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL133553; CAC36090.1; -.
DR HSP; P04004; LOC0.
DR InterPro; IPR000585; Hemopexin.
DR InterPro; IPR001212; Somatomedin_B.
DR Pfam; PF00045; Hemopexin; 2.
DR Pfam; PF01033; Somatomedin_B; 2.
DR PRINTS; PR00022; SOMATOMEDINB.
DR SMART; SM00120; HX; 2.
DR SMART; SM00201; SO; 2.
DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN 1.
DR PROSITE; PS00524; SOMATOMEDIN B; 2.
SQ SEQUENCE 1404 AA, 151076 MW, 782A11746B3FDEE5 CRC64;

Query Match      99.4%; Score 5797.9; DB 2; Length 1404;
Best Local Similarity 96.1%; Pred. No. 1.1e-102;
Matches 1096; Conservative 0; Mismatches 3; Indels 41; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQOVSSQ----- 25
DB 1 MAWKTLPIYLLLLSVFVIQOVSSQDLSSCAGRCGEGYSRDATCNCYHMECCPDF 60
QY 26 -----ELSCKGRCFESFERGECDDAOCKKYDKCCPDYEFSCAEVHNPTSPSSKKAP 79
DB 61 KRVCCTAELSCKGRCFESFERGECDDAOCKKYDKCCPDYEFSCAEVHNPTSPSSKKAP 120
QY 80 PPSGASQTIKSTTKRSPKPPNKKTKVIESEITEHSVSENQESSSSSSSSSSSTIW 139
DB 121 PPSGASQTIKSTTKRSPKPPNKKTKVIESEITEHSVSENQESSSSSSSSSSSTIR 180
QY 140 KIKSSKNSAANRELQKLVKONKNRTKKPTPKPPVVDAGSLDNGDFKVTTPDTST 199
DB 181 KIKSSKNSAANRELQKLVKONKNRTKKPTPKPPVVDAGSLDNGDFKVTTPDTST 240
QY 200 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSILVNKETTIVETKETTINKQSTSDG 259
DB 241 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSILVNKETTIVETKETTINKQSTSDG 300
QY 260 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTTPKEPAS 319
DB 301 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTTPKEPAS 360
QY 320 TTPKEPTTTPKSAPTTPKEPATTTKSAPTTPKEPATTTKBPATTTKBPATTTKBP 379
DB 361 TTPKEPTTTPKSAPTTPKEPATTTKSAPTTPKEPATTTKBPATTTKBPATTTKBP 420
QY 380 APTTTSAPTTTPKEPATTTTPKBPATTTTPKBPATTTTPKBPATTTTPKBPATTPK 439
DB 421 APTTTSAPTTTPKEPATTTTPKBPATTTTPKBPATTTTPKBPATTTTPKBPATTPK 480
```







```

RL Cytogenet. Cell Genet. 90:291-297(2000).
DR EMEL; AB034730; BAA92310.1; -.
DR HSSP; P04004; 10C0.
DR MGD; MGI:1891344; PrG4.
DE InterPro; IPR000585; Hemopexin.
DE InterPro; IPR001212; Somatomedin_B.
DR Pfam; PF00045; Hemopexin; 2.
DE PRINTS; PR00022; Somatomedin_B; 2.
DR SMART; SM00120; HX; 2.
DR SMART; SM00201; SO; 2.
DE PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR PROSITE; PS00524; SOMATOMEDIN_B; 2.
SQ SEQUENCE 1054 AA; 115991 MW; 4FC64BFA42282335 CRC64;

Query Match      44.9%; Score 2622; DB 2; Length 1054;
Best Local Similarity 46.7%; Pred. No. 1.8e-42;
Matches 541; Conservative 41; Mismatches 147; Indels 430; Gaps 23;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25
Db 1 MGWKILPVLCLLLPVLLIQVSSQDLSSCAGRCGEGYSRDATCNDYNCQHYMECCPDF 60
QY 26 -----ELSCGRCFESPERGECDCDAOCKYDKCCPDYSEFCABVHNPTSPSSKKAP 79
Db 61 KRVCSPFLSCGRCFESFARGRECDQSQCKQKCCADYDFCEVHNSTSPSKTAP 119
QY 80 PPSGASQTIKSTTKRSPKPNKKTKKVIESTHEHSVSNQESSSSSSSSSSSTTW 139
Db 120 TPAGASDITKSTTKRSPKSPST-TRTIKVVESELTTEHSDSENQESSSSSSSSSS--TIR 176
QY 140 KTKSKNSANRELQKLVKONKQRTKKTPPKPPVVDEAGSLDNGDPKVT--TPPT 197
Db 177 KTKSKNS-ANRELQKNPNVKNKNTPKPKPNPEPPAVDEAGSLDNGEFLTPPPDP 235
QY 198 STQHNKVTSPKITTAKPINRPSLPNPSDSKETSLSVWNETTVETKETTNTKQTS 257
Db 236 PTPHSKVAISPKTTAAKVPVTPKPSLAPNSETSKASLASNKETTIVETKETAINKQSSA 295
QY 258 DGKEKTSKAKETQSIEKTSKADLAPTSKVLAKTPKAETTTKGPALTTKPEPTTPPKPF 317
Db 296 -SKKKTTSVKETSAEKTSKDVETS----- 321
QY 318 ASTPKPEPTTKISAPTTKPEAPTTTKSAPTTKPEAPTTTKSEAPTT-----PK 369
Db 322 -----TPKNSAPTTTKPEVTTTKESKFLPLPQ 349
QY 370 EPAPTTKBPAPTTKSAPTTKPEAPTTKPKAPTTKPEAPTTKPEPTTPPKPEAPT 429
Db 350 EPBPTAKGPPPTTKKPEPTTKPEPTTKPEPTTKPEPTTKPEPTTKPEPTTKPEPT 409
QY 430 TKBPAPTTKPEAPTTAKKPAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTTKSA 489
Db 410 TKKPEPTTKPEGPTTKPEPTTKPEPTTKPEPTTKPEPTTKPEPTTKPEPTTKPE 461
QY 490 PTTTKBPAPTTKSAPTTKPEPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPE 549
Db 462 -----EPTTKPEPTTKPEPTTKPEPTTKPEPTTKPEPTTKPEPTTKPEPTTK 507
QY 550 EPAPTTTKBPAPTTAKKPAPTTKPEAPTTKPKLTTPTPKLAPTTKPEAPTTPELAP 609
Db 508 EPBPT----- 512
QY 610 TTPBPTTKPEAPTTKPKAAAPNTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPE 669
Db 513 -----TPKPEPTTKPEPTTKPEPTTKPEPTTKPEPTTKPEPTTKPEPTTKPE 538
QY 670 GTAPTTLKBPAPTTKPKAPKELAPTTKPEPTTKPEPTTKPEPTTKPEAPTTKPE 729
Db 539 -----KEPEPTTKPKP-----EPTTKPEPTTKPEPTTKPEPTTKPEPTTK 562
QY 730 KEPAPTTKGTAPTTLKBPAPTTKPKAPKELAPTTTKGPSTSTSDKPAPTTKPEAPT 789

```

## RESULT 6

```

Q9VR49
ID Q9VR49 PRELIMINARY; PRT; 1225 AA.
AC Q9VR49;
DT 01-WAY-2000 (Tremblrel. 13, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE CG3047-PA.
GN Name=Sgsl; ORFNames=CG3047;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan X.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Brokstein P., Brotter P.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brokstein P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann M.,
RA Glöckner A., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Gosler C., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Markulov G., Malsbina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M., M.G.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

```



QV 1073 FMPEVTPDMXYLPRVFNQ 1090  
 Db 1150 ---PVVPQWPMGPYNQ 1164

## RESULT 7

MUC2\_HUMAN  
 ID MUC2\_HUMAN STANDARD; PRT; 5179 AA.  
 AC Q02817; Q14878;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Mucin 2 precursor (Intestinal mucin 2).  
 GN Name=MUC2; Synonyms=SMUC;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Intestine;  
 RX MEDLINE=94132002; PubMed=8300571;  
 RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;  
 RT "Molecular cloning of human intestinal mucin (MUC2) cDNA."  
 RI Identification of the amino terminus and overall sequence similarity  
 RI to prepro-von Willebrand factor.";  
 RL J. Biol. Chem. 269:2440-2446(1994).  
 [2]  
 RP SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.  
 RC TISSUE=Colon;  
 RX MEDLINE=93016075; PubMed=1400449;  
 RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,  
 RA Kim Y.S.;  
 RT "The human MUC2 intestinal mucin has cysteine-rich subdomains located  
 RT both upstream and downstream of its central repetitive region.";  
 RL J. Biol. Chem. 267:21375-21383(1992).  
 [3]  
 RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.  
 RX MEDLINE=91358717; PubMed=1885763;  
 RA Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,  
 RA Petersen G.M., Kim Y.S.;  
 RT "MUC-2 human small intestinal mucin gene structure. Repeated arrays  
 RT and polymorphism.";  
 RL J. Clin. Invest. 88:1005-1013(1991).  
 CC -!- FUNCTION: Coats the epithelia of the intestines, airways, and  
 CC other mucus membrane-containing organs. Thought to provide a  
 CC protective, lubricating barrier against particles and infectious  
 CC agents at mucosal surfaces.  
 CC -!- SUBUNIT: Multimeric.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: Colon, small intestine, colonic tumors,  
 CC bronchus, cervix and gall bladder.  
 CC -!- PTM: All cysteine residues are involved in intrachain or  
 CC interchain disulfide bonds (By similarity).  
 CC -!- POLYMORPHISM: The number of repeats is highly polymorphic and  
 CC varies among different alleles.  
 CC -!- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.  
 CC -!- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.  
 CC -!- SIMILARITY: Contains 2 WVFC domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; L21998; AA095295.1; -;  
 CC EMBL; M74027; AAA59875.1; -;  
 CC EMBL; M94131; AAA59163.1; -;  
 CC EMBL; M94132; AAA59164.1; -;

PIR: A49963; A43932.  
 Genew; HGNC:7512; MUC2.  
 MIM; 158370; -;  
 DR InterPro; IPR002919; Cysrich\_TIL.  
 DR InterPro; IPR006208; Cys\_knot.  
 DR InterPro; IPR006207; Cys\_knot\_C.  
 DR InterPro; IPR006209; EGF\_like-  
 DR InterPro; IPR001007; VWF\_C.  
 DR InterPro; IPR001846; VWF\_D.  
 DR Pfam; PF00007; Cys\_knot; 1.  
 DR Pfam; PF01826; TIL; 1.  
 DR Pfam; PF00093; VWC; 1.  
 DR Pfam; PF00094; VWD; 4.  
 DR SMART; SM00041; CT; 1.  
 DR SMART; SM00214; VWC; 2.  
 DR SMART; SM00216; VWD; 4.  
 DR PROSITE; PS01185; CTCK\_1; 1.  
 DR PROSITE; PS01225; CTCK\_2; 1.  
 DR PROSITE; PS00022; EGF\_1; UNKNOWN\_1.  
 DR PROSITE; PS01308; VWF\_1; 2.  
 DR PROSITE; PS0184; VWF\_2; 2.  
 KW Glycoprotein; Repeat; Signal.  
 FT SIGNAL 1 20 Potential.  
 FT CHAIN 21 5179 Mucin 2.  
 FT DOMAIN 1401 1747 APPROXIMATE REPEATS.  
 FT REPEAT 1401 1416 1.  
 FT REPEAT 1417 1432 2.  
 FT REPEAT 1433 1448 3.  
 FT REPEAT 1449 1464 4.  
 FT REPEAT 1465 1471 5.  
 FT REPEAT 1472 1478 6.  
 FT REPEAT 1479 1494 7A.  
 FT REPEAT 1495 1517 7B.  
 FT REPEAT 1518 1533 8A.  
 FT REPEAT 1534 1556 8B.  
 FT REPEAT 1557 1572 9A.  
 FT REPEAT 1573 1596 9B.  
 FT REPEAT 1597 1612 10A.  
 FT REPEAT 1613 1635 10B.  
 FT REPEAT 1636 1651 11A.  
 FT REPEAT 1652 1675 11B.  
 FT REPEAT 1676 1683 12.  
 FT REPEAT 1684 1699 13.  
 FT REPEAT 1700 1715 14.  
 FT REPEAT 1716 1731 15.  
 FT REPEAT 1732 1747 16.  
 FT DOMAIN 4815 4886 WVFC 1.  
 FT DOMAIN 4924 4991 WVFC 2.  
 FT DOMAIN 5075 5160 CTCK.  
 FT DISULFID 5075 5122 By similarity.  
 FT DISULFID 5089 5136 By similarity.  
 FT DISULFID 5098 5152 By similarity.  
 FT DISULFID 5102 5154 By similarity.  
 FT DISULFID ? 5159 By similarity.  
 FT CARBOHYD 163 163 N-linked (GlcNAc...)  
 FT CARBOHYD 423 423 N-linked (GlcNAc...)  
 FT CARBOHYD 670 670 N-linked (GlcNAc...)  
 FT CARBOHYD 770 770 N-linked (GlcNAc...)  
 FT CARBOHYD 894 894 N-linked (GlcNAc...)  
 FT CARBOHYD 1139 1139 N-linked (GlcNAc...)  
 FT CARBOHYD 1154 1154 N-linked (GlcNAc...)  
 FT CARBOHYD 1215 1215 N-linked (GlcNAc...)  
 FT CARBOHYD 1230 1230 N-linked (GlcNAc...)  
 FT CARBOHYD 1246 1246 N-linked (GlcNAc...)  
 FT CARBOHYD 1787 1787 N-linked (GlcNAc...)  
 FT CARBOHYD 1820 1820 N-linked (GlcNAc...)  
 FT CARBOHYD 4339 4339 N-linked (GlcNAc...)  
 FT CARBOHYD 4351 4351 N-linked (GlcNAc...)  
 FT CARBOHYD 4362 4362 N-linked (GlcNAc...)  
 FT CARBOHYD 4373 4373 N-linked (GlcNAc...)  
 FT CARBOHYD 4422 4422 N-linked (GlcNAc...)  
 FT CARBOHYD 4438 4438 N-linked (GlcNAc...)  
 FT CARBOHYD 4502 4502 N-linked (GlcNAc...)

FT	CARBOHYD	4616	4616	N-linked (GlcNAc. . .)	(Potential).
FT	CARBOHYD	4627	4627	N-linked (GlcNAc. . .)	(Potential).
FT	CARBOHYD	4752	4752	N-linked (GlcNAc. . .)	(Potential).
FT	CARBOHYD	4787	4787	N-linked (GlcNAc. . .)	(Potential).
FT	CARBOHYD	4881	4881	N-linked (GlcNAc. . .)	(Potential).
FT	CARBOHYD	4888	4888	N-linked (GlcNAc. . .)	(Potential).
FT	CARBOHYD	4955	4955	N-linked (GlcNAc. . .)	(Potential).
FT	CARBOHYD	4970	4970	N-linked (GlcNAc. . .)	(Potential).
FT	CARBOHYD	5019	5019	N-linked (GlcNAc. . .)	(Potential).
FT	CARBOHYD	5038	5038	N-linked (GlcNAc. . .)	(Potential).
FT	CARBOHYD	5069	5069	N-linked (GlcNAc. . .)	(Potential).
FT	CONFLICT	1351	1351	H -> L (in Ref. 3).	
FT	CONFLICT	1412	1412	T -> S (in Ref. 3).	
FT	CONFLICT	1449	1449	L -> P (in Ref. 3).	
FT	CONFLICT	1504	1504	M -> T (in Ref. 3).	
FT	CONFLICT	4192	4192	G -> S (in Ref. 2).	
SQ	SEQUENCE	5179 AA; 540295 MW; 85CD7571FB9A5663	CRC64;		

Query Match.  
Best Local Similarity 24.3%; Score 1354.4; DB 1; Length 5179;  
Matches 431; Conservative 70; Mismatches 478; Indels 796; Gaps 66;

QY	2	AWKTLPIY-----	-----LLLLLSVFI	-----QQVSQE 26
Db	1020	SWKEAPCPDVSNPECSLNPHRRSWAEKQCSILKSSVFSICHSKYDVPKPFVEACVHDS	1079	
QY	27	LSCK--GRCFESFERGECDDA-----	---OCKKYDKC-----	CPDYSEFCAEHNPT 70
Db	1080	CSDCTGGDC-----	---ECFSAVASYAQCTKEGACVFWRTPLCP---	IFC-DYNNPP 1127
QY	71	SP-----	-----PSSKKAPP-----	80
Db	1128	HECEWHYPCGNRSFETCRTINGIHNSISVYLEGCVPRCKDRPIVEEDLKKCVTADCK	1187	
QY	81	-----PSGAS-----	---QTIKS-----	90
Db	1188	GCVEDTHYPGASVPTPECKSCVCTNSSQVCRPEEGKILNQDGAFCYWEICGPNG	1247	
QY	91	-----	---TKRSPKPNKKTKKVISSSEITTEBHSVSENEQSSSS 129	
Db	1248	TVEKHFNICSLITRPSLTITFTITLPTTISFTT	-----TTT 1285	
QY	130	SSSSSSSTIIWIKSKNSAANRELQKCLKVDKNKQNTKKKTPKPWW-----	DEAG 182	
Db	1286	TTTPTSTGV-----	-----LSTTPKCLWSMDINEDHPS 1315	
QY	183	SGLDNGDFK-----	---VTTPDTSTTOH-NKVSTSPK-----	210
Db	1316	SGSDGDGDRPEFDVCGGAPEDIKCRSKNDPHLSLEQHGQKVCQDSVSGVICKNBDQFGNGP	1375	
QY	211	-----	---ITTAKPINRPSSLPPNSDTSKETSILTWNKETTIVETKETTT 250	
Db	1376	FGLCYDYKIRVNCWPMQKCIITTPSPPTTTPSPPTTITLPLTTPTPSPPTT	-----TTT 1430	
QY	251	TNKQSTDGKEKTSIAKETQSIBKTSAKOIAPTSKVLAKPTPKAEATTTKGPAITTPKEPT	310	
Db	1431	TPPPTTPSPPIITTTTP-----	---LPIT-----TPSPPISTITTPPTTPSPPT 1474	
QY	311	-----PTTKEPASITPKPTPTTIKSAK-TPKPEP-----	APTTKSAK-TPKPEP 356	
Db	1475	TPSPPTTTPSPPTTITTPPTTTPSPPTTTPSPMTTITTPASITLTPPTTPSPPTTITTPP	1534	
QY	357	PTTTKEPAPTTPKEP-----	APTTKPAPITTKKSAK-TPTKEPAPTTPKPAITTP 407	
Db	1535	PTTTPSPPTTPIPTPSTITLPLPTTTPSPPTTITTPPTTTPSPPTTTPSPPTTITTP	1594	
QY	408	KEPAPTTPKEPTTTPKPEAPTTPKEAPTTPKEP-----	APTAPKPAITTPKPEAPT 460	
Db	1595	TPPPTTPSPPTTITTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT	1654	
QY	461	TPKEPAPTTPKEPPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAK-----	TPTKPEPSTTK 517	
Db	1655	TPSPPTTTPSPPTTITTPPTTTPPTTTPSPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTT	1714	





QY 475 -----PTTPKEPA-----PTTKSA-----PT----- 491  
Db 605 SKTTTHETAPATKKTTHETPTQKSTTLRIETPTTKSSTARTTPTTKRETTERT 664  
QY 492 -----TTKEPA-----PTTKSA-----PTTKPEP 511  
Db 665 QEPSTKTTTHETAPATKKTTHETPTQKSTTLRIETPTTKSSTAKTTTRETRET 724  
QY 512 SPPTTKPEA-----PTTPK-----EPA-----PT 530  
Db 725 TERTTKEPTTKTTTHETPTTKKTTHETPTTKSTTLKTEPTTKRKTSTTKTTEPT 784  
QY 531 TPKKAPATTKPEA-----PTTPKEPATTKKPA-----PTAPK-----EPA 568  
Db 785 TKRKTERTTKEPTTKTTTHETPTTKSTTLRIETPTTKSSTTLKTEPT 840  
QY 569 -----PTTPKETA-----PTTPK-----PTTPKLT----- 584  
Db 841 TRKSTTKTTEPTTKRKTTHETPTTKKTTHETPTTKSTTLKTEPTTKRKTSTTKTTEPT 900  
QY 585 ---PTT-PEKLAPTPEKAPTPEELAPTTPEPT-----PTTPE-----EPA- 624  
Db 901 TEPPTTKSSTAKTTRE---PTTKRETTERTKPTTKTTTHETPTTKTTEPTTKTTEPTT 957  
QY 625 -----PTTPKAAENTPKEPAPTTPKEPATTPKEPA-----PTTPK 661  
Db 958 KKSSTLKPTEPTTKRKTSTTKTTR---PTTKRKTERTTKEPTTKTTTHETPTTKT 1015  
QY 662 ETA-----PTTPKGTA-----PTTLKEPATTPKPAKELA-----PTTKPTS 702  
Db 1016 TTTTKTTTHETTKSTLKEPTTEPTTKKTSTTKTTEPTTKRVTTERTTPTTKTTTH 1075  
QY 703 TTSKPEA-----PTTPKGTA-----PTTPKEPATTPKEPA-----PTTP 737  
Db 1076 ETTAEPATKTTHETPTQKSTTLRIETPTTKSSTAKTTREPTTKRETTERTTKEPTT 1135  
QY 738 KGAPTTLKEPA-----PTTPK-----KPAPKELAPTTTKGPTSTSDKPAPTPKETA 786  
Db 1136 KTTTHKTTEPTTKTTTHETPTTKSTTLKPTES---PTTRK---TSITTKTTEPTTKRKT 1190  
QY 787 PTPKPEPA-----PTTPK-----KP-----APTPEP 808  
Db 1191 ERTKPTTKTTTHKTTEPTTKTTTKTTTHETPTTKSTTLKTEPTTKRKTSTTKT 1250  
QY 809 PPTTSEVSTPTTKP-----TTIHKSPDE-----STPBLSEPTPKALESPKPGVP 858  
Db 1251 REPTTKRVTTERTTPTTKTTTHKTTEPTTKTTTKTTTHETPTTKKTTTHETPTTK-STITLKTEEP 1309  
QY 859 TTKTPAATKPEMTTAKKTTERDLR-----TTTPE 888  
Db 1310 TTKTSTTKTTEPTTKRVTTERTTPTTKTTTHKTTEPTTKTTTKTTTHETPTTK 1369  
QY 889 TTTAAP-----KWKETA-----TTTKTESKITATTTQVTSITTTQDTTFFKIT 933  
Db 1370 STTLKTEPTTKRKTSTTKTTEPTTKRVTTERTTPTTKTTTHKTTEPTTKKT 1427  
QY 934 TLKT-----TLAP-----KVTTK-----KTIITTEIMNKP-----EE 962  
Db 1428 TKKTTHETPTTKSTTLKTEPTTKRKTSTTKTTEPTTKRVTTERTTPTTKTTTHKT 1487  
QY 963 TAKPKDRANSKATTPKQ-----KP-----TKAPKKT----- 991  
Db 1488 TEPPTTKTTTKTTTHETPTTKSTTLKTEPTTKRKTSTTKATREPTTKRVTTERTTREP 1547  
QY 992 -----STPKPKTMRVRKPKTTPTRKKTSTMPELNP 1023  
Db 1548 TTRKTTAKTTTEPTTKTTTKTTTHETPTTKSTTL-----KTEPTTKRKTSTTKTTRP 1603  
QY 1024 TTRIAEMLQTT-----TRPNQTPNSKLVEVNPXSEDAGAEGET----- 1063  
Db 1604 TTRTSTSVKTTADQTKRRTAEMSTNQEPSTVEITTNSSQNSQNTTSTTTEBQVHH 1663

QY 1064 -----PHMLLRPH----- 1071  
Db 1664 HHHHHYHKPADLGPSTILPDLPLPLPLWPPLPLPEILPLPLPLTALPLPLPL 1723  
QY 1072 -----VFMEVPT-PDMYDLPRVPN 1089  
Db 1724 PPLPEVNLTALSLPEISLPLNPLPPLQPLPN 1752  
RESULT 9  
AAS64673  
ID AAS64673 PRELIMINARY; PRT; 1761 AA.  
AC AAS64673;  
DT 01-APR-2004 (TRMBLrel. 27, Created)  
DT 01-APR-2004 (TRMBLrel. 27, Last sequence update)  
DT 01-APR-2004 (TRMBLrel. 27, Last annotation update)  
DE CG33300-PA (Fragment).  
GN CG33300.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Ananides P.G., Scherer S.E., Li P.W., Hoekins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,  
RA Abail J.P., Aghayani A., An H.J., Andrews-Pfankoch C., Baldwin D.,  
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fostler K.J., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostalko M., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,  
RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Furl V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kianos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zeng X.H., Zhong F.N., Zhong W., Zhou S., Zhu S., Zhu H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster";  
RL Science 287:2185-2195(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22426065; PubMed=12537568;  
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,  
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,  
RA Paclab J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
RA Weinstein G., Scherer S.E., Myers S.E., Gibbs R.A., Rubin G.M.;







QY 725 -----APTP-----KEPAPTPKGTATTLKEPAP----- 750  
Db 2128 TTTTPTTRTSTDSTWSSMSASTPEPSTTGTTRTTRTPTDSTWSSMSASTTP 2187  
QY 751 -----TPKPPAPKELA-----PTTKGPTSTISDKAPT-----TPK 783  
Db 2188 GTTPTPTPTSTESTDITWSSASTPEPSTTGTTRTTRTPTDSTWSSMSASTPE 2247  
QY 784 E-----TAPTTPKEPAP-----TTPKKPAP----- 803  
Db 2248 PSTTPTGTRTTRTPTDITWSSMSASTPEPSTTGTTRTTRTPTDSTWSSMS 2307  
QY 804 -----TPE----- 807  
Db 2308 SASTPESTTPELLOQDHLPIQCRKLVCRQPLLRHRLVQPELLQDHLPIQCRQ 2367  
QY 808 -----TPPTT-----SE 815  
Db 2368 PLLRHLRLVQPELLQDHLPIQCRKLVCRQPLLRHRLVQPELLQDHLPIQCRQ 2427  
QY 816 VSTPTTKPTTHKSP-----DESTPELSAEP----- 843  
Db 2428 ASTPESTTGTTRTTRTPTSTESTDITWSSASTPEPSTTGTTRTTRTPTSTESTD 2487  
QY 844 -----TPKALENSKPE-----GVPTTKTAATKPEMTTAKOKTTERDRTTP 887  
Db 2488 TWSSASTPEPSTTGTTRTTRTPTSTESTDITWSSASTPEPSTTGTTRTTRTPT 2545  
QY 888 ETTTAAPKMTKATTT-----TEKTTESKITATT 916  
Db 2546 TDTTMSASTPEPSTTGTTRTTRTPTDITWSSASTPEPSTTGTTRTTRTPTST 2605  
QY 917 TQVSTTTQDTPFKITTKITLAP-----KVTIT 947  
Db 2606 TTMSSMSASTPEPSTTGTTRTTRTPTDITWSSMSASTPEPSTTGTTRTTRTPT 2665  
QY 948 KKTITITTEIMNKPEETAKPDATN-----SKATTPKQ----- 981  
Db 2666 ESTDITWSSASTPEPSTTGTTRTTRTPTDITWSSMSASTPEPSTTGTTRTPT 2725  
QY 982 KPT-----KAPKPTSTK-----KPKTMP-----RVRKP 1005  
Db 2726 RPTPTDITWSSMSASTTGTTRTTRTPTSTESTDITWSSASTPEPSTTGTTRTPT 2785  
QY 1006 KTTPTPRKMTSTM-----PELNP-TSRIAEMLOTTTRPNTPNKSLVENVKPEDAG 1057  
Db 2786 RPTPTDITWSSMSASTPEPSTTGTTR-----TTPTRPTDITWSSMSAST----- 2835  
QY 1058 GAEGTTPHMLLRHVFMPEVTP 1079  
Db 2836 -----TPE-----PSTTP 2843

RESULT 11  
C6SSE6  
ID Q6SSE6 PRELIMINARY; PRT; 3409 AA.  
AC Q6SSE6;  
IT 05-JUL-2004 (T-EMBLrel. 27, Created)  
LT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)  
IT 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)  
LE plus agglutinin.  
CN Name=SAGI;  
CS Chlamydomonas reinhardtii.  
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
CC Chlamydomonadaceae; Chlamydomonadaceae.  
CX NCBI\_TaxID=3055;  
EN [1]  
FP SEQUENCE FROM N.A.  
FA Ferris P.J., Waffenschmidt S., Umen J.G., Ishida K., Kubo T., Lau J.,  
EA Goodenough U.W.;  
EL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY450930; AAS07044.1; -

DR InterPro; IPR002951; Atrophin.  
DR InterPro; IPR003882; Pistil\_extensin.  
DR PRINTS; PRO1222; ATROPHIN.  
DR PRINTS; PRO1218; PSTILEXTENSIN.  
SQ SEQUENCE 3409 AA; 336045 MW; 7FE87633EDD6631F CRC64;  
Query Match 20.3%; Score 1185.8; DB 2; Length 3409;  
Best Local Similarity 15.1%; Pred. No. 1.5e-14;  
Matches 345; Conservative 143; Mismatches 411; Indels 1382; Gaps 60;  
QY 1 MAWKTLPI-----YLLLLLSV----- 16  
Db 50 LAWVTIPVSDALQFPVNVSVTSSAATDAPSAIAYIGALLDLVNWVPQNCINDTRRSYD 109  
QY 17 ----- 16  
Db 110 APWPSRCALPAVQGYGYDETYIDATVYSGNSLRPFYSTCRYPASSDPYAFLEPYSMPWD 169  
QY 17 -----FVIQVSSQ-----ELSCGR-----CFE 35  
Db 170 DRGIQPVLDGDMWGFVNVVDSNLKWIDFAWTAQAGAWLGQDFRDVWVNGFYHCSW 229  
QY 36 SFERGECDCDAQCKYD----- 53  
Db 230 PFV-----EC-----SSCENYDIADPYDPDKIANGIVPAVITALDPENASLYIYILDGFMFPAG 283  
QY 54 -----KCCPD----- 58  
Db 284 SLDNVWYLNAYNFIGGPLEANLPTLLPSLOHALDHCRAITPDVGRGTASLOYGWOQYPS 343  
QY 59 --YE----- 60  
Db 344 GQPEYECSDVGTGSDGTYEIVSGMTPDEWGDVAAAASDSSLPWANLRTVRLSNQALY 403  
QY 61 -----SFCAEVN----- 68  
Db 404 PIPEGLRSASSISSWRLQNTLQGLPEFAAPINSLLYLGLTKLGTHTVWHDPAHQ 463  
QY 69 -----PTSPSS-----SKAPPPSGASQTI-----KST 91  
Db 464 GGECLAPPFPSPSPRPPLPPLPSPPLPPLPSPFPVPPSPSPSPSPSPSPSPSPSP 523  
QY 92 TKRSPPPNKKTKKVIIESEITEE-----H 117  
Db 524 PELPPSPFPPTVARGIQVGGICDPSMPSPRPSPSPSPSPSPSPSPSPSPSPSPSP 583  
QY 118 SVSENGESSSSSSSSSSSTIWKIKSSKNSAANRELQKKLVKDNKNRTKKKTPKPV 177  
Db 584 PPSPDSPSPASS-----VPPSPPEP----- 602  
QY 178 VDEAGSGLDNGDFKVTTPTDSTTQHNKVSISPKITTAKPINPRPSLPNSDTSKETS 237  
Db 603 -----SPKPPSPAPPSPAPPSP----- 620  
QY 238 NKETTIVETKETTNTKQSTDGKKTSAKETQSIKTSKDLAPTAKVLAKTPKAEIT 297  
Db 621 -----PPSPAPPSPAPSP----- 633  
QY 298 TKGPAITTPKEPTPTTPKEPAST-----TPKEPTTIKSAPTTPKEPAPTTKSA 350  
Db 634 ---PAPPSPQPPSPFPVPPQPPSPFPSPFPSPFPSPFPSPFPSPFPSPFPSPFP 690  
QY 351 TPKEPAPTTKEPAPTTK-----EPAPTTKEPAPTTKSAPTTPKEPAPTTKKA 405  
Db 691 SPAPPLPSPFPSPFPSPFPSPFPSPFPSPFPSPFPSPFPSPFPSPFPSPFPSP 750  
QY 406 -----TPKEPAPTTKEPAPTTTPKEPAPTTKEPAPTTTPKEPAPTT 448  
Db 751 PSPEPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP 810  
QY 449 PAPTTPKEPAPTTTPKEPAPT-----TTKEPSPTPK 479  
Db 811 PEPSPSPAPPSPFPSPFPSPFPSPFPSPFPSPFPSPFPSPFPSPFPSPFPSP 870



D0	404	P	EGLRSASSSSISWRLOQNTLCLGPLPEFAAPINSLLYLGTKLTHTVWHDPAH7Q	463
QY	69	-	- - - - -PTSps- - - - -SKKAPPSPGASQTl- - - - -KGT	91
D0	464	G	GECLAPPSPSPRRPPRLPSPPPPLLPSPFPVPSPSPSPSPSPSPSPSPSP	523
QY	92	T	KRSPKPNNKKTKKVIESEITEE- - - - -H	117
D0	524	P	LPPSPSPSPTPVARCIVGGICDPSMPSPSPRPQPSPSPSPSPSPSPSPSPSP	583
QY	118	S	VENQBSSSSSSSSSSSTIWIKISKNSAANRELQKLKVKONKNRTKKXPTKPPV	177
D0	584	P	PSFDSPPAS- - - - -VPSSEPP- - - - -602	
QY	178	V	EAGSGLDNGDFKVTTPDTSTTQHNVKVS7PKITTAKPINPRPSLPNPNSDTSKETSLTV	237
D0	603	-	- - - - -SPKPPSFAPSPAPPSP- - - - -620	
QY	238	N	KETTVEKTNTTNKQSTDGKEKTTSAKETOSIEKTSAKOLAPT8KVIAKTPKAETT	297
D0	621	-	- - - - -PSPAPSPAPS- - - - -633	
QY	298	T	KPALTTKPEPTTPKEAST- - - - -TPKEPTTIKSAPTTKPEAPTTTKSAPT	350
D0	634	-	- PAPSPQPFVFVPQPSPPSPKPPSPAPPSPVPSPAPPSPAPPSPAPPSPAPP	690
QY	351	T	KEPAPTTTKEAPATPK- - - EPAPTTTKEAPTTTKSAPTTKPEAPTTTKKCAPT-	405
D0	691	S	PAPPLPSPSPAPPSPSPSPAPPSPAPSPAPPSPAPSPAPPSPAPPSPAPPSP	750
QY	406	-	- - - - -TPKEPATTTKPEPTTTTKPEAPTTKPEAPTTKPEAPTT- - - - -APKK	448
D0	751	P	SPSPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP	810
QY	449	P	APTTKPEAPATTTPKEAPT- - - - -TTKPSSTTPK	479
D0	811	P	EPSPPSPAPPSPSPSPAPPSPGFPPSQPPSPSPSPAPPSPVPSPAPPSPSPSE	870
QY	480	E	APTTTKSAPTTTKEAPTTTKSAPTTKBPSTTTTKEAPTTKPEAP-	529
D0	871	P	SPSPSPAPPSPSPSPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP	929
QY	530	-	- - - - -TTPKXPAPTTKPEAPTTTKEAPTTTKKPA7APK- - - - -565	
D0	930	P	PPSPAPPSPAPPSPSPSPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPP	989
QY	566	-	- - - - -EPAPTTKETA7TTTKL7TTP7EKLA7TTP7- - - - -KPA7TTP7ELAP7-	610
D0	990	S	POPPSPAPSPAPPSPSNPSNPSPVPTTPSP-GPPSPSPSPAPSPPTTPTTSPPPPP	1048
QY	611	-	- - - - -TPBEPTTPEEAPTTTEKAAFN- - - - -TPKEAPTTTKEAPATPK- - - - -653	
D0	1049	P	SPSPSPAPPPLPSPSPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP	1108
QY	654	E	PAPTTTKETA7TTPKOTAP7TL- - - - -K	677
D0	1109	S	PAPSPSPAPPSPSPSPAPLLPSPDPSPAPPSPMPPLP7SPSPSPVPPTPPPS	1168
QY	678	E	PAPTTTKKAPKELAP7T- - - - -TKETSTSUKPAP7TTP	713
D0	1169	P	PAPSPAPPSPQPLEPSPSPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPP	1228
QY	714	K	GAPTTTKEAPTTTKEAPTTTG7AT7TLK7EAP7TTPK7KAPKELAP7-T7KG7TST	772
D0	1229	A	PPSP-APPSPAPPSPAPPSPAPSPOPPS- - - - -PAPSPSPSPAPPSPAPPSPAPPSP	1283
QY	773	T	SUKPAPTTTKEAPTTTKEAPTTTKKAP7TTPETTPPT-TSPSVST7TTP7TKE7THKS	831
D0	1284	A	PPSPAPPSP--APPSPSPAPPQPPSPVPSPAPPSPPTTPAPAPAALPPLPSPSPA	1340
QY	832	P	DESTPELSABPTKALE- - - - -NSPKDPGV- - - - -PTTK7P	863
D0	1341	P	FLVPVPASPAPSPFLRPPQFTAMPPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPP	1400

[illegible]

RESULT 13	
SLP1_CLOTM	
ID SLP1_CLOTM	STANDARD; PRT; 1664 AA.
AC	Q06852;
DT	01-JUN-1994 (Rel. 29, Created)
DT	01-JUN-1994 (Rel. 29, Last sequence update)
DT	05-JUL-2004 (Rel. 44, Last annotation update)
DE	Cell surface glycoprotein 1 precursor (Outer layer protein B) (S-layer protein 1).
DE	protein 1.
GN	Name=olp8;
OS	Clostridium thermocellum.
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC	Clostridium.
OX	NCBI_TaxID=1515;
RN	[1]

SEQUENCE FROM N.A.  
 RC STRAIN=NCIB 10682;  
 RX MEDLINE=93209931; Pubmed=8459832;  
 RA Fujino T., Beguin P., Aubert J.-P.;  
 RT "Organization of a Clostridium thermocellum gene cluster encoding the  
 RT cellulosomal scaffolding protein Cipa and a protein possibly involved  
 RT in attachment of the cellulosome to the cell surface.";  
 RL J. Bacteriol. 175:1891-1899(1993).  
 CC -/- SUBUNIT: Assembled into mono-layered crystalline arrays.  
 CC -/- SUBCELLULAR LOCATION: Cell wall.  
 CC -/- SIMILARITY: Contains 4 S-layer homology (SLH) domains.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X67506; CAA47841.1; -;  
 DR PIR; T18262; T18262;  
 DR InterPro; IPR008965; Cellul\_bind.  
 DR InterPro; IPR001119; SLH.  
 DR Pfam; PF00335; SLH; 3.  
 DR PROSITE; PS01072; SLH\_DOMAIN; 2.  
 KW Cell wall; Repeat; S-Layer; Signal.  
 FT SIGNAL 1 28 Potential.  
 FT CHAIN 29 1664 Cell surface glycoprotein 1.  
 FT DOMAIN 36 763 4 X 156 AA approximate repeats.  
 FT REPEAT 36 191 1.  
 FT REPEAT 207 363 2.  
 FT REPEAT 409 565 3.  
 FT REPEAT 607 763 4.  
 FT DOMAIN 771 1377 approximate tandem repeats of T-P-S-D-E-P.  
 FT  
 FT DOMAIN 1378 1449 Gly/Pro/Ser/Thr-rich.  
 FT DOMAIN 1453 1494 SLH 1 (incomplete).  
 FT DOMAIN 1495 1565 SLH 2.  
 FT DOMAIN 1566 1625 SLH 3.  
 FT DOMAIN 1626 1646 SLH 4 (incomplete).  
 FT SEQUENCE 1664 AA; 178194 MW; 5F396695BA9FE74B CRC64;  
 SQ  
 Query Match 19.6%; Score 1143.3; DB 1; Length 1664;  
 Best Local Similarity 21.2%; Pred. No. 4e-14;  
 Matches 367; Conservative 131; Mismatches 425; Indels 807; Gaps 66;  
 QY 4 KTLPIVILLLLSVFV-----IQVSSQE 26  
 DB 6 KVLISILLTLLIISTTVSNMSFAEATPSIEMVLDKTEVHVGDVITATIKVNNIRKLAGYQ 65  
 QY 27 LSCGRCFBSFERGECDCDAQCKYDKCPDYSECAEVHNTPSPSSK----KAPPS 82  
 DB 66 LNIK-----FDPEVLQVDPATGEEFTDKSMP-- 92  
 QY 83 GASQTIKTKRSPKPNKKTKKVIESEITEHSHVSENOESSSSSSSSSSSTTWKIK 142  
 DB 93 -VNRVLLTNSKYGTP-----VAGNDIKSGIINFATGNNLTAYK 131  
 QY 143 SS----- 144  
 DB 132 SSGIDEHTGIIGIFKVLKKQNTSIRFEDTSLMPGALSGTSLFDWDAETITGYEVIQPD 191  
 QY 145 -----KGAANRELQK-KLVKD-----NKKN-----RTKKKP----- 172  
 DB 192 LIVVEAPLKASVALELDKTKVKGDIITAIKIKENKNFAGYQINIKYDPTMLEAIEL 251  
 QY 173 -----PKPVVDSEAGSLDN----- 187  
 DB 252 ETGSAIAKRTWPTGTVQLSDNYGKTTAVANDVGAGIINFABAYSNLTKYRTGVAET 311  
 QY 188 -----GPKVTTPTD-----STQHNKV----- 205

Db QY 312 GIICKIGFRVLKAGSTAIRFEDTTAMPGAIEGTYPMDWYGENIKGYSVQGEIVAESEE 371  
 QY 206 -----STSPKITTAKPINRPSLP-----PNS-----DTSK----- 231  
 Db 372 PGSEPTPEPVTETPVDPTPTVTEEPVPSLPDSYVIMELDKTKVKVGDIIATIKIENM 431  
 QY 232 -----ETSLTVNKETTVETKETTTTKTKQSTDKGKETS----- 265  
 Db 432 KNFAGYQLNLIKDYPTMLEALETGSIAKRTWPVTGGTV-----LOSDNYGKTTAVAND 486  
 QY 266 -----AKETQSIEKTSADLAPTSKVLAKEPTKPAETTT----- 298  
 Db 487 VGAGIINFABAYSNLTKYRTGVAETGIIICKIGFRVLKAGSTAI-----RFEDTTAMPG 541  
 QY 299 -----KGPALVTP-----KEPTPT-TPKEPASTTPKEPTPT----- 328  
 Db 542 AIBGTYMFMDWYGENIKGYSVQGEIVAEGEETPEPVTETPVDPTPTVTEEPVPSLP 601  
 QY 329 ----- 328  
 Db 602 DSVVIMELDKTKVKEGDVIIATIRVNNIKNLGAYQIGIKYDKPVLEAFNIETGDPIDEGT 661  
 QY 329 ----- 329  
 Db 662 WPAVGGTILKNRDYLPVGVAINNVSKILNFAAYVYVDFDYREEGKSDTGIIIGNIGFRV 721  
 QY 330 IKSAPTPK-----EPAPTTKSAPTPKBPAPPT 359  
 Db 722 LKAEDTTIRFEELSMPGSDGTMYMLDWNLNRSYVVIQAPAKAAS-----DEIPLTD 776  
 QY 360 TKEPAPTTKPEPAP-----TTTKEPAPTTKSAPTTKPEPAPTTKPKPAPTTKPEPAPTT 413  
 Db 777 TPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPEEPPTPTPTPSDEPTPSDEPTPSDE 836  
 QY 414 -----TPK-EPTPT-TPKEPAPT-TKEPAPTTKPEPAPTAPKKA-----PTTKPEPAPT- 460  
 Db 837 PTPSDEPTPSDEPTPSDEPTPEEPPTPTPTPSDEPTPSDEPTPSDEPTPSDEPTPSDE 896  
 QY 461 TPKEPAPTTTKPEPAPTTKPEPAPTTKSAPTTKKBPAPTTKSAPTTKKBPAPTTKBPAP 520  
 Db 897 TPEEPITPTPSDEPTPSDEPTPS-----DEPTPSDEPTPSDEPTPSDEPTPEEPITPTD 953  
 QY 521 PTTKPEPAPTTKKA-----EPTTKPEPAPT-TPKEPAPTTKBPAPTTKBPAPTTAPK 565  
 Db 954 PTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPEEPPTPTPTPSDEPTPSD 1013  
 QY 566 EPAPTTPKETAPTTPKKLTPTTPEKLAPTTTPKEPAPT-----TPBELAPTTPEEPPT-TPE 621  
 Db 1014 EPTPSD-----EPTPSDEPTPSDEPTPSD-EPTPSDEPTPSDEPTPE 1052  
 QY 622 EPAPTTPKAAAPNTPKEPAPTTKPEA-----PTTKPEPAPT-TPKEPAPTTKGTAPTTLKE 678  
 Db 1053 EPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPEEPITPTDPSDEPTPSDE 1112  
 QY 679 PAPTTPKAPKAPKELAPTTTKPT-STTSKAPAPTTKGTAPTTKBPAPTTKBPAPTTKBPAPTT 737  
 Db 1113 PTPS-----DEPTPSD-EPTPSDEPTPSDEPTPEEPITPTDPSDEPTPSDEPTPS-----DEPTPS-- 1165  
 QY 738 KGTAPTTLKEPAPT-TPKKAPKAPKELAPTTTKGPTSTTSKAPAPT-----TPKETAPTTKPEP 793  
 Db 1166 --DEPTPSDEPTPSDEPTPEEP-----PTDTPSDEPTPSDEPTPSDEPTPSD-EPTPSDEP 1217  
 QY 794 APT-TPKKAPAPTTPEPTTSEVSTPTTKPTTHKSPDSTPBLSPAEPTPKALENSP 852  
 Db 1218 TPESTPEEPITDTPSDEPTPSD-EPTPSDEPT-----PSDEPTP--SDEPTP--SETP 1266  
 QY 853 KEPGVPTTKIPAATKPEMTTAKDKITRDLKTTTETTTAAKPKTKETATTTTKEKTESKI 912  
 Db 1267 EEP-IPDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPEEP 1325  
 QY 913 TATTQVSTTTQDPTTKKTLTKTLAKVTTTKKTLTTTMMKPBETAKPKDRATN 972  
 Db 1326 PTDTPSDEPTPSDEPTPSDEPT-----PSDEPTPSDEPT 1360

```

QY 973 SKATTPKQ-KTKAPKPTSTKKXTMPRVKPKTTPTRKMTS----- 1016
DQ 1361 SDEPTSEDEPTSEPT-----PTTPTPTSTPTSGSGSGGGG 1407
QY 1017 -----TMPELNPTSRLAEAMLQTTTRPNOTPNKSLVENPKSBDAGAGETPHMLL 1068
DQ 1408 GGGGTVPTPTPTIS-----KPTSTPAPTEIE-BPTSDVGAIGGEHAYL 1455
QY 1069 RPV-----VFMP----- 1076
DQ 1456 RGYPDGSPERNITRAEAAVIFAKLLGADESYGASPSYDLADTHWAAWAIKPAFSQ 1515
QY 1077 -----VTPD-----MDYLPRVFNQGI-----IINP 1096
DQ 1516 GLFKGYDGTFRPDQNIIRAFAVVLHFLTKVKGQEIIMSKLATIDISNP 1565

RESULT 14
Q3WQ4 PRELIMINARY; PRT; 1349 AA.
AC Q8WQ4;
DI 01-MAR-2002 (TrEMBLrel. 20, Created)
DI 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DI 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Mucin 5 (Fragment).
GN Name=MUC5AC;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RV [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21426417; PubMed=11535137;
RA Escande F., Aubert J.P., Porchet N., Buisine M.P.;
RT "Human mucin gene MUC5AC: organization of its 5'-region and central
RT repetitive region.";
RL Biochem. J. 358:763-772(2001).
DR EMBL; AJ298318; CAC83675.1; -.
FT NON TER 1
FT NON TER 1349 1349
SQ SEQUENCE 1349 AA; 135599 MW; 4DC3C1544F1E5EBA CRC64;

Query Match
Best Local Similarity 19.1%; Score 1117.1; DB 2; Length 1349;
Matches 355; Conservative 73; Mismatches 371; Indels 709; Gaps 44;

QY 74 SSKKAPPSPGASQTIKSTTKRSPKPKPKKKKKVIESEITEHVSSENOESSSSSSSS 133
DQ 3 STTSAPTSTTSTSTST-----STPQTSTTSASTS 35
QY 134 SSGTIWKSSKNSAANRELQKKLVKDNKNRTKKKP--TPKPPVVDEAGSLDNGDFK 191
DQ 36 ITS-----GPGTTPSP----- 46
QY 192 VITPDTSTQHKNKVSIPKITTAKPINRPSLPNSDTSKETSILVNKETTIVTKETTTT 251
DQ 47 --VPTTSTT-----SAPTTSTSAATSTTSAPTSTTSAPTST 83
QY 252 NKQTSIDGKEK-----TTSAKETQSEKTSKADLAP-----TSKVLA 288
DQ 84 TTSASTASKTSGLGTTPSPPIPTTSTSPPTTSTASTASKTSGPGTTPSPVTTSTIPA 143
QY 289 KPTPKAPT-----TTKGALTPPKPTPTT----- 313
DQ 144 ---PRSTTSASTSTTPGPG-ITP-SPVPTTSTASVSKTSTSHVSKITTHSQPVTRDC 198
QY 314 -----PKE----- 316
DQ 199 HLRCTWTKWFDVDFPPSPGPHGGDKETNNIIRSGEKICRRPEITRLQRAKSHPEVSIE 258
QY 317 -----PASTTP----- 322

```

```

DQ 259 HLGQVVQCSREBGLVCRNQDQGPFKMCLNVEVRVLCCETPKGCPTVSTPTVAPSTPSGK 318
QY 323 -KEPT-----PTTIKSAPTTKEPAPTT-TKSAPTTKEPAPTTTKEPAP 365
DQ 319 ATSPTQSTSSWOKSRITTLVTSTSTPQTSTTSAPTSTTTSAPTSTTSPT 378
QY 366 TTPKEPAPTTTKEPAPTTK-----SAPTTKEPAPTTTKEPAPTTTKEPAPTTK 416
DQ 379 QTSISSAPTSTTSAPTSTTSAPTSTTSAPTSTTSAPTSTTSAPTSTTSAPTST 438
QY 417 EPTPTTKEPAP-----TTKEPAPTT----- 437
DQ 439 TSTPQTSKTSAASTSTSSGTTTPSPVTTTSTASVSKTSTSHVSVSKTTHSQPVTRDCHP 498
QY 438 -----PKE----- 440
DQ 499 RCTWTKWFDVDFPPSPGPHGGDKETNNIIRSGEKICRRPEITRLQRAKSHPEVSIEHL 558
QY 441 -----PAPTAKKP 449
DQ 559 GQVVQCSREBGLVCRNQDQGPFKMCLNVEVRVLCCETPKGCPTVSTSVTASPLVGEPP 618
QY 450 APT-----TPKEPAPTTKEPA--PTTKPEPSPTTKEPAP 483
DQ 619 AQTQSTSSWOKSRITTLVTSSITSTTQSTSTTSAPTSTTTPASIPSTTS--APTSTTSAP 676
QY 484 TT-TKSAPTTTKEPAP-TTTKSAPTTKEPSPTTKEPAPTTTKEPAPTTTKEPAPTTK 541
DQ 677 TTTSTTSAPTSTTTPQTSTTSAPTSTTSAPTSTTSAPTSTTSAPTSTTSAPTST 736
QY 542 EPAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 601
DQ 737 TSAPTSTSS-APTNTTSAPTSTTSAPTSTTSAPTSTTSAPTSTTSAPTSTTPTTP 795
QY 602 TPEELAPTTPEEPTTPEEAPTTPKAAAPNTKEPAPTTKEPAPTTKEPAPTTKEPAPTTK 661
DQ 796 QSTSTSPSTTSAPTSTTTSAPTSTTSTTQSTTSAPTSTTSAPTSTTSAPTSTTSAPTST 855
QY 662 ETAPITPKGTAPTTLKEPAPTTK-----KPAPKELAPTTTKEPTSTTS-- 705
DQ 856 TSFHTTSTSPSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 912
QY 706 -----DKAP----- 710
DQ 913 HVSVKITHSQPVTRDCHPRCTWKFVDVDFSPGPHGGDKETNNIIRSGEKICRRPEE 972
QY 711 -----TTPKG 715
DQ 973 ITRLOCAEASHPEVSIHLGQVVQCSREBGLVCRNQDQGPFKMCLNVEVRVLCCETPKG 1032
QY 716 -----TAPTPK-----EPAPTTKEPA 733
DQ 1033 CPVTSTPTVAPSTPSGRATSPQSTSSWOKSRITTLVTSTSTPTSTPTSTTSAPTSTIPA 1092
QY 734 PTPPKGTAPTTLKEPAPTTKPKAPKELAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 793
DQ 1093 STPSTTSAPTSTTSAPTSTTSAPTSTTSAPTSTTSAPTSTTSAPTSTTSAPTST 1147
QY 794 APPTPKKAPTTPEPTPTTSEVSTPT--TTKEPTTIHKSPPDESTPELSAEPPTPKALENS 851
DQ 1148 APPTSTISASTSTISAPTSTTSAPTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 1202
QY 852 PKEPGVPTTKTAAATKPEMTTAKOKTTERDLRTPPETTTAAAPKMTKEATITTEKTESK 911
DQ 1203 P-----VPTTSTTSAPTSTTSAPTSTT-----SGFGTTPSPVPVPSITSAATSTTSAP 1253
QY 912 ITATTTQVSTTTQ--DTPPKITTLTKTTLAPKVTTTKKTTITTEIMNKPEETAKPKOR 969
DQ 1254 TTRTTSAPTSSMTSGGTTTPSPVPTTSTTS-APTST----- 1289
QY 970 ATNSKATTPKPKQKPTKAPKKP-TSTKKPKTMPRVKPKTTPTRKMTSTMPELNPTSRIA 1028
DQ 1290 -TSGPGTTPSPVPTTSTTSAPTST-----TSGFGTTPSPVPVPTTSTSA--PTTSTT 1338

```

```

QY 1029 EAMLOTTT 1036
DB 1339 SASTATT 1346

RESULT 15
Q8IR51
ID Q8IR51 PRELIMINARY; PRT; 3432 AA.
AC Q8IR51;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE CG32602-PB.
GN ORFNames=CG32602;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos K., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Fertaz C., Ferreira S., Fleischmann W.,
RA Fessler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasse K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhang M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RN Science 287:2185-2195 (2000).
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverly T., Murny D.M., Nelson C.R.,
RA Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila
RT melanogaster euchromatic genome sequence.";
RN Genome Biol. 3:RESEARCH0079-RESEARCH0079 (2002).

```

```

RN RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatic
RT a genomic perspective.";
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084 (2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review.";
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083 (2002).
RN [5]
RP SEQUENCE FROM N.A.
RX FLYBASE;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RX FLYBASE;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
RN EMBL; AE003495; AA09587.1; -
DR FLYBase; Fg00052602; CG32602.
SQ SEQUENCE 3432 AA; 349492 MW; FE907AC75578657A CRC64;

Query Match 19.1%; Score 1112.3; DB 2; Length 3432;
Best Local Similarity 26.0%; Pred. No. 3.9e-13;
Matches 466; Conservative 68; Mismatches 410; Indels 847; Gaps 92;

QY 56 CPDYEFCAEVHNTSP-----SSSSSSSTIWK-----IKSSKNSAAN 150
DB 58 CPSPHTTPACEHRITSPATPHCHPHATPHNCGHDTSSSRATTPKPTLETTIGTTFK 117
QY 73 -----PSSKAPPSPGAS---QIKSTKSPKPKPKKTKKVKVIESEITEHSVSE 121
DB 118 TTQITTRGTTQK--PTEGTSAPOTLKTEGTAKPTLKPT-----EGTAKPTTLK 169
QY 122 NOESSSS-----SSSSSSSTIWK-----IKSSKNSAAN 150
DB 170 PTEGTSAPOTLKTEGTAKPTLKTEGTAKPTLKTEGTAKPTLKTEGTAKPTLKTEGTTA- 228
QY 151 RELQKKLVKDNKNKTKKPKPKPVVDAGSLDNGDFKVTTPDTS----- 198
DB 229 ----KPTTLKPTK--GTTAKPTTLK-----TEGTSAPOTLKTEGT 265
QY 199 ----TTOHNVKSTSPKITTAKP-----INPRSLPPNSDTSKETSLSL----- 235
DB 266 TAKPTTLKTEGTSAPOTLKTEGTAKPTLKTEGTAKPTLKTEGTAKPTLKTEGTSAPOTLK 325
QY 236 ----TVNKETTVEKETT-----TTNKQTS-----TDGK 260
DB 326 PTEGTTAKPTLKTEGTSAPOTLKTEGTAKPTLKTEGTSAPOTLKTEGTSAPOTLKTEGTS 385
QY 261 EKT-----TSAKET--QSIKETSATK--DLAPTSKVLAKPT---PKAETTKGPALTTPK 307
DB 386 PTLKTEGTSAPOTLKTEGTSAPOTLKTEGTSAPOTLKTEGTSAPOTLKTEGTSAPOTLK 444
QY 308 EPT---PPTTKPEASTTPK---EPT-----PTTIK-----SAPTTKEPAPTTKSA 348
DB 445 EGTSAKPTTLKTEGTSAPOTLKTEGTAKPTLKTEGTAKPTLKTEGTSAPOTLKTEGTTAK-- 502
QY 349 PTPPK-----EPAPTTTKEPA-----PTTKPEPAPTTTKEPAPTT-- 383
DB 503 PTLKTEGTSAPOTLKTEGTAKPTLKTEGTSAPOTLKTEGTSAPOTLKTEGTTAK---PTTLK 559

```



```
QY 384 ----TKSAPTKEPAPTPK---KPA-----PTTPKEPAPTPK-----BPTPTPK 425
Db 560 PTEGTSAPKPTTLKPTTEGTTAKPTTLKPTTEGTSAPKPTTLKPTTEGTTAKPTTLKPTTEGTSAPK 619
QY 426 PA-----PTTKEPA-----PTTPKEPAPTPKPAKPTTPK 455
Db 620 PTTLKPTKGTAKPTTLKPTTEGTSAPKPTTLKPTDGTAKPTTLK---PTEGTSAPKPTTLK 676
QY 456 EPAPTPPK-----BPAPTTKEPS-----PTTPKEPAPTPPK 487
Db 677 PTEGTTAKPTTLKPTTEGTSAPKPTTLKPTDGTAKPTTLKLTTEGTSAPKPTTLKPTTEGTTAK 736
QY 488 -----SAPTTTKEPA-----PTT-----TKSAPTTPKESPPTTKEPA-----520
Db 737 PTTLKPTTEGTSAPKPTTLKPTTEGTTAKPTTLKPTTEGTSAPKPTTLK-PTKGTAKPTTLKPT 795
QY 521 -----PTTPKEPAPTPPK-----KPA-----PTTPKEPAPTPPK--549
Db 796 EGTSAPKPTTLKPTTEGTTAKPTTLKPTTEGTSAPKPTTLKPTTEGTTAKPTTLKPTTEGTTAKPT 855
QY 550 ---EPAPTTTKKPA---PTAPKEPAPTPKETAPTTPKKLT-----PTT-----587
Db 856 TLKPTDGTAKPTTLKPTTEGTSAPKPTTLKPTTEGTTAKPTTLKPTTEGTTAKPTTLKPTDGT 915
QY 588 ---PEKLAPTTPKEPAPTT-----PEELAPT--TPPEPTPTT 619
Db 916 TAKPTTLKPTTEGTSAPKPTTLKPTDGTAKPTTLKPTTEGTSAPKPTTLKPTTEGTTAKPTTLK 975
QY 620 PEE---PAPTPPK-----AAANPTKEPAPTPPKBAPTPPK-----EPA-----656
Db 976 PTEGTTAKPTTLKPTTEGTSAPKPTTLKPTTEGTTAK---PTTLKPTTEGTSAPKPTTLKPTTEG 1032
QY 657 ---PTTPKETAPTTPKGT-----PTTLKEPAPTPPK---KPA-----PKELA 693
Db 1033 SAKPTTLKPTDGTAKPTTLKPTTEGTSAPKPTTLKPTTEGTTAKPTTLKPTTEGTSAPKPTTLK 1092
QY 694 PT-----TTKEPTSTSDKPA-----709
Db 1093 PTEGTSAPKPTTLKPTTEGTTAKPTTLKPTTEGTSAPKPTTLKPTTEGTTAKPTTLKPTTEGTSAPK 1152
QY 710 PTPPKGTAPTTPKEPAPTPPKBAPTPPKGTA-----PTTLKEPAPTPPK---K 755
Db 1153 PTTLKPTTEGTTAK---PTTLKPTTEGTTAKPTTLKPTTEGTSAPKPTTLKPTTEGTTAKPTTLK 1209
QY 756 PA-----PKELAPT-----TTKGTSTSDKPA-----778
Db 1210 PTEGTSAPKPTTLKPTTEGTTAKPTTLKPTTEGTSAPKPTTLKPTTEGTTAKPTTLKPTDGTAK 1269
QY 779 -----PTTPKETAPTTPKEPAPTPPK-----KPA-----PTT--805
Db 1270 PTTLKPTTEGTSAPKPTTLKPTTEGTTAK---PTTLKPTTEGTTAKPTTLKPTDGTAKPTTLK 1326
QY 806 -----PETPPPTTSVSTPT-----TTKEPTTIHKSPDESTP 837
Db 1327 PTEGTSAPKPTTLKPTTEGTTAKPTTLKPTTEGTSAPKPTTLKPTDGTAKPTTL---KPTEGT-1383
QY 838 ELSABPT-----PKALENSPKE-----PGVPTTKIPAAKPEMTTAKD 876
Db 1384 --SAKPTTLKPTTEGTTAKPTTLK---PTEGTSAPKPTTLKPTTEGTTAKPTTLKPTTEGTTAKP 1439
QY 877 KTTTERDLRTT-----PETTTAAPKMTKETATTTTEKTESKITAT 915
Db 1440 TTTLKPTTEGTTAKPTTLKPTTEGTSAPKPTTLKPTTEGTTAKPTTLKPTTEGTTAKPTTLKPTKG 1499
QY 916 T-----916
Db 1500 TSAKPTTLKPTTEGTTAKPTTLKPTDGTAKPTTLKPTTEGTSAPKPTTLKPTTEGTTAKPTTL 1559
QY 917 -----TQVTS-----TTQDPTT--PFKITTLKTTTLAP-----942
Db 1560 KPTEGTSAPKPTTLKPTTEGTSAPKPTTLKPTTEGTSAPKPTTLKPTTEGTTAKPTTLKPTTEG 1619
```

```
QY 943 KYTTTKKLTITTTIMNKPEETAKPKDRATNSKATTEKPOKPTKAPKKPT-----STKK 995
Db 1620 KPPTTLKPTTEGTTA---KP--TTLKPTD-GTAKPTTLKPTTEGTSAPKPTTLKPTTEGTTAK 1672
QY 996 PKTMPRVVRKP-----KITTPTRKMTSTMP-ELNPTSRIAEAMLOTTTRPNQTPNSKLV 1047
Db 1673 PTTL---KPTEGTSAPKPTTLKPTTEGTSAPKPTTLKPTTEGTSAPKPTTLKPTTEGTTAKPT 1726
QY 1048 EYNPKSSEDAGGABGET--PHMLLRPHVFMPEVTPDMDYLPVRPNQGIINP 1096
Db 1727 TLKP-----TEGTTAKPTTL-----KPTEGTTAKP 1751
```

Search completed: October 13, 2004, 12:16:27  
Job time : 151.835 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 18.2347 Seconds  
(without alignments)  
3815.116 Million cell updates/sec

T-tle: SEQ1-D

Perfect score: 5576

Sequence: 1 MAWKTLPIYLLLLSVFIQ.....DMDYLPVNOGIINPMLS 1049

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.1

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued Patents AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A-COMB.pdp.\*  
2: /cgn2\_6/ptodata/1/iaa/5B-COMB.pdp.\*  
3: /cgn2\_6/ptodata/1/iaa/6A-COMB.pdp.\*  
4: /cgn2\_6/ptodata/1/iaa/6B-COMB.pdp.\*  
5: /cgn2\_6/ptodata/1/iaa/PCFUS-COMB.pdp.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pdp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5576	100.0	1049	4	US-07-757-022B-58
2	5576	100.0	1313	4	US-07-757-022B-142
3	5561.9	99.7	1354	4	US-07-757-022B-48
4	5561.9	99.7	1363	4	US-07-757-022B-52
5	5546.9	99.5	1140	4	US-07-757-022B-104
6	5546.9	99.5	1404	4	US-07-757-022B-2
7	5546.9	99.5	1404	4	US-07-757-022B-62
8	5546.9	99.5	1404	4	US-09-298-970A-1
9	5523.9	99.1	1404	4	US-10-164-595-78
10	5547.9	99.7	1314	4	US-07-757-022B-50
11	5393.3	96.7	1320	4	US-07-757-022B-46
12	5393.3	96.7	1320	4	US-07-757-022B-60
13	5384.3	96.6	1320	4	US-10-164-595-58
14	5379.2	96.5	1361	4	US-07-757-022B-40
15	5358.7	96.1	1038	4	US-07-757-022B-74
16	5358.7	96.1	1270	4	US-07-757-022B-44
17	5344.6	95.9	1311	4	US-07-757-022B-42
18	5235.9	93.9	1022	4	US-07-757-022B-84
19	5011	89.9	941	4	US-07-757-022B-14
20	1346.8	24.2	5179	4	US-09-538-092-1258
21	1088.5	19.5	8991	4	US-08-714-741-32
22	827.1	14.8	3256	4	US-09-919-172-98
23	827.1	14.8	3256	4	US-09-919-172-98
24	827.1	14.8	3256	4	US-09-919-172-98
25	816.3	14.6	3118	3	US-09-579-181-1
26	809.6	14.5	2972	3	US-09-579-181-2
27	762.9	13.7	1837	3	US-08-928-361B-5

28 762.9 13.7 1837 4 US-09-588-995A-5  
29 756.1 13.6 2142 4 US-09-538-092-1142  
30 735 13.2 1721 3 US-08-700-651-5  
31 733 13.1 1721 3 US-08-928-361B-6  
32 733 13.1 1721 4 US-09-588-995A-6  
33 716 12.8 2736 4 US-09-252-991A-30227  
34 705.3 12.6 807 4 US-09-294-663-4  
35 697.5 12.5 4019 4 US-08-854-133-425  
36 692.7 12.4 3969 3 US-08-061-376-5  
37 692.6 12.4 2468 4 US-09-976-594-726  
38 692.6 12.4 2468 4 US-09-538-092-1135  
39 687.4 12.3 3969 4 US-09-538-092-1262  
40 686.1 12.3 805 3 US-09-103-429A-4  
41 679.9 12.2 788 4 US-09-294-663-3  
42 675 12.1 827 4 US-09-248-796A-17307  
43 659.8 11.8 786 3 US-09-103-429A-3  
44 649.9 11.7 220 4 US-07-757-022B-96  
45 640.2 11.5 207 4 US-07-757-022B-116

## ALIGNMENTS

RESULT 1  
US-07-757-022B-58  
; Sequence 58, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION NUMBER:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cseri, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 58:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1049 amino acids  
; TYPE: AMINO ACID

**This Page Blank (uspto)**

```
Qy 121 PTPKPPVDEAGSLDNGDKVTTPTDSTTOHNVKSTSPKITTAKPINRPSLPPNSDTS 180
Db 121 PTPKPPVDEAGSLDNGDKVTTPTDSTTOHNVKSTSPKITTAKPINRPSLPPNSDTS 180
Qy 181 KETSLTVNKETTVEKETTITNKQTSDDGKEKTTSAKETOSIEKTSKAKOLAPTSKVLAKP 240
Db 181 KETSLTVNKETTVEKETTITNKQTSDDGKEKTTSAKETOSIEKTSKAKOLAPTSKVLAKP 240
Qy 241 TPXAETTTKGPAUTTPKBPPTTPKBPASTTPKEPTPTTIKSAPPTPKBPAPTTKSAPT 300
Db 241 TPXAETTTKGPAUTTPKBPPTTPKBPASTTPKEPTPTTIKSAPPTPKBPAPTTKSAPT 300
Qy 301 TPKEPAPTTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTT 360
Db 301 TPKEPAPTTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTT 360
Qy 361 APPTKEPTPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTT 420
Db 361 APPTKEPTPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTT 420
Qy 421 KPSPTTPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTT 480
Db 421 KPSPTTPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTT 480
Qy 481 TPKKPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTT 540
Db 481 TPKKPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTT 540
Qy 541 LAPTTKEPAPTTPEELAPTTPEEPTPTTPEEAPTTPKAAAPNTPKBPAPTTKBPAPTTKBPAPTT 600
Db 541 LAPTTKEPAPTTPEELAPTTPEEPTPTTPEEAPTTPKAAAPNTPKBPAPTTKBPAPTTKBPAPTT 600
Qy 601 TPKEPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTT 660
Db 601 TPKEPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTT 660
Qy 661 TTPKGTAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTT 720
Db 661 TTPKGTAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTT 720
Qy 721 STTSKAPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTT 780
Db 721 STTSKAPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTT 780
Qy 781 SPDESTPELSAETPKALENSKBPVPTTKTBAATKPEMTTTAKOKTTERDLRTTPETT 840
Db 781 SPDESTPELSAETPKALENSKBPVPTTKTBAATKPEMTTTAKOKTTERDLRTTPETT 840
Qy 841 TAAPKMTKETAATTTKTESKITATTTQVNTTODTTPKIIITLKTTLAPKVVTTTKT 900
Db 841 TAAPKMTKETAATTTKTESKITATTTQVNTTODTTPKIIITLKTTLAPKVVTTTKT 900
Qy 901 ITTATEIMNKPEETAAPKADRAATNSKATTPKPKPTSKAPKPTSTSKPKTMRVRKPKTTPT 960
Db 901 ITTATEIMNKPEETAAPKADRAATNSKATTPKPKPTSKAPKPTSTSKPKTMRVRKPKTTPT 960
Qy 961 PRKWTSTMPNLNPTSRIAEAMLOTTTPRPNOTPNKSLVEVNPKSEDAGAGETPHMLLRP 1020
Db 961 PRKWTSTMPNLNPTSRIAEAMLOTTTPRPNOTPNKSLVEVNPKSEDAGAGETPHMLLRP 1020
Qy 1021 HVFMPEVTPDMXYLPRVFNQGIINPMLS 1049
Db 1021 HVFMPEVTPDMXYLPRVFNQGIINPMLS 1049
```

## RESULT 3

US-07-757-022B-48

Sequence 48, Application US/07757022B

Patent No. 6433142

GENERAL INFORMATION:

APPLICANT: Gesner, Thomas G.

APPLICANT: Clark, Stephen C.

```
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESS: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseri, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 1354 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-48
```

```
Query Match 99.7%; Score 5561.9; DB 4; Length 1354;
Best Local Similarity 96.2%; Pred. No. 1.3e-169;
Matches 1049; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

Qy 1 MAWKTLPIYLLLLSVFVIQQVSSQ----- 25
Db 1 MAWKTLPIYLLLLSVFVIQQVSSQDLSACAGRCGEGYSDATCNCNDYQHMECCPDF 60
Qy 26 -----ELSCGKCFSEFERGECDDAQCCKYDKCCPDYESFCABEHSVSENOSSSS 79
Db 61 KRVTAELSCKGKCFSEFERGECDDAQCCKYDKCCPDYESFCABEHSVSENOSSSS 120
Qy 80 SSSSSSTTWIKSSKNSAANRELQKLVKDNKNKRTKKKPTKPPVVDAGSLDNGD 139
Db 121 SSSSSSTTWIKSSKNSAANRELQKLVKDNKNKRTKKKPTKPPVVDAGSLDNGD 180
Qy 140 FKVTTPDSTTOHNVKSTSPKITTAKPINRPSLPPNSDTSKETSITVNKETTVEKETT 199
Db 181 FKVTTPDSTTOHNVKSTSPKITTAKPINRPSLPPNSDTSKETSITVNKETTVEKETT 240
Qy 200 TTNKQTSDDGKEKTTSAKETOSIEKTSKAKOLAPTSKVLAKPTKAEITTKGPAUTTPKEP 259
Db 241 TTNKQTSDDGKEKTTSAKETOSIEKTSKAKOLAPTSKVLAKPTKAEITTKGPAUTTPKEP 300
Qy 260 TPTTPKEPASTTPKEPTPTTIKSAPPTPKBPAPTTKBPAPTTKBPAPTTKBPAPTTPK 319
```

```

Db 301 TPATPKPEASTPKPEPTTTIKSAPTTPKEPAPTITTKSAPTTPKEPAPTITTKEPAPTTPK 360
QY 320 EPAPTITTKEPAPTITTKSAPTTPKEPAPTITTKKAPAPTTPKEPAPTITTKPEPTTPKEPAPT 379
Db 361 EPAPTITTKEPAPTITTKSAPTTPKEPAPTITTKKAPAPTTPKEPAPTITTKPEPTTPKEPAPT 420
QY 380 TKEPAPTTPKEPAPTAPKAPAPTTPKEPAPTITTKPEPAPTITTKESPTTPKEPAPTITTKSA 439
Db 421 TKEPAPTTPKEPAPTAPKAPAPTTPKEPAPTITTKPEPAPTITTKESPTTPKEPAPTITTKSA 480
QY 440 PTTTKEPAPTITTKSAPTTPKEPSPTTTTKPEPAPTITTKPEPAPTITTKKAPAPTTPKEPAPTTPK 499
Db 481 PTTTKEPAPTITTKSAPTTPKEPSPTTTTKPEPAPTITTKPEPAPTITTKKAPAPTTPKEPAPTTPK 540
QY 500 EPAPTITTKKAPAPKAPAPTTPKETAPTTPKLTTPPKLAPTTPPEKLAPTTPPEKAPTTPESLAP 559
Db 541 EPAPTITTKKAPAPKAPAPTTPKETAPTTPKLTTPPKLAPTTPPEKLAPTTPPEKAPTTPESLAP 600
QY 560 TTPEEPTPTTPEEPAPTTPKAAAPNTPKPEPAPTITTKPEPAPTITTKPEPAPTITTKETAPTTPK 619
Db 601 TTPEEPTPTTPEEPAPTTPKAAAPNTPKPEPAPTITTKPEPAPTITTKPEPAPTITTKETAPTTPK 660
QY 620 GTAPTTLKEPAPTTPKAPKAPKALAPTTPKSTSDKAPAPTTPKGTAPTTPKGTAPTTPK 679
Db 661 GTAPTTLKEPAPTTPKAPKAPKALAPTTPKSTSDKAPAPTTPKGTAPTTPKGTAPTTPK 720
QY 680 KEAPTTPKGTAPTTLKEPAPTTPKAPKAPKALAPTTPKSTSDKAPAPTTPKGTAPTTPKGTAPTTP 739
Db 721 KEAPTTPKGTAPTTLKEPAPTTPKAPKAPKALAPTTPKSTSDKAPAPTTPKGTAPTTPKGTAPTTP 780
QY 740 PKEPAPTTPKAPAPTTPETPEPTTSEVSTPTTKEPTTIHKSPDESTPELSABETPKALE 799
Db 781 PKEPAPTTPKAPAPTTPETPEPTTSEVSTPTTKEPTTIHKSPDESTPELSABETPKALE 840
QY 800 NSPKGPGVPTTKPAATKPMWTTAKDITTERDLRTPTTETTTAAPKMTKETATTTTEKITE 859
Db 841 NSPKGPGVPTTKPAATKPMWTTAKDITTERDLRTPTTETTTAAPKMTKETATTTTEKITE 900
QY 860 SKITATTTQVSTTTQDTPFKITLTKTTLAPKVVTTTKITTTTEIMNKPEETAKPKDR 919
Db 901 SKITATTTQVSTTTQDTPFKITLTKTTLAPKVVTTTKITTTTEIMNKPEETAKPKDR 960
QY 920 ATNSKATTPKPKPTKAPKPTSTTKPKTTPRVKPKPTTTPPKWTSTMPELNPTSRIAE 979
Db 961 ATNSKATTPKPKPTKAPKPTSTTKPKTTPRVKPKPTTTPPKWTSTMPELNPTSRIAE 1020
QY 980 AMLQTTTRPNQTPNSKLVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPVPN 1039
Db 1021 AMLQTTTRPNQTPNSKLVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPVPN 1080
QY 1040 QGIIINPMLS 1049
Db 1081 QGIIINPMLS 1090

```

## RESULT 4

```

US-07-757-022B-52
; Sequence 52, Application US/07757022B
; Patent No. 6433142

```

## GENERAL INFORMATION:

```

; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseerr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)878-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 1363 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-52

```

```

Query Match 99.7%; Score 5561; DB 4; Length 1363;
Best Local Similarity 95.5%; Pred. No. 1.4e-169;
Matches 1049; Conservative 0; Mismatches 0; Indels 50; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQELSCKRCFCFESFERGECDCDAQCKKYDKCCPDYE 60
Db 1 MAWKTLPIYLLLLSVFVIQVSSQELSCKRCFCFESFERGECDCDAQCKKYDKCCPDYE 60
QY 61 SFCA-----EHSVS 70
Db 61 SFCAEHNPTSPSSKAPPPSGASQTIKTTKRSPKPNKTKKVIIESEEITEHSVS 120
QY 71 ENQESSSSSSSSSSSTIWKIKSSKNSAANRELQKKLVKDNKNRTKKKPTPKPPVUDE 130
Db 121 ENQESSSSSSSSSSSTIWKIKSSKNSAANRELQKKLVKDNKNRTKKKPTPKPPVUDE 180
QY 131 AGSGLDNGDFKVTTPDSTTQHNVKSTSPKITTAKPINRPSLPSPNSDTSKETSILTYNKE 190
Db 181 AGSGLDNGDFKVTTPDSTTQHNVKSTSPKITTAKPINRPSLPSPNSDTSKETSILTYNKE 240
QY 191 TTVETKETTTNKQSTDGKEKTTSAKETQSIKTSKADLAPTSKVLAKTPKAETTTKG 250
Db 241 TTVETKETTTNKQSTDGKEKTTSAKETQSIKTSKADLAPTSKVLAKTPKAETTTKG 300
QY 251 PALTTPEPTTPPKPEASTTPKEPTPTTIKSAPTTPKEPAPTITTKSAPTTPKEPAPTITTK 310
Db 301 PALTTPEPTTPPKPEASTTPKEPTPTTIKSAPTTPKEPAPTITTKSAPTTPKEPAPTITTK 360
QY 311 KEPAPTTPKEPAPTITTKPEPAPTITTKSAPTTPKEPAPTITTKKAPAPTTPKEPAPTTPKEPTP 370
Db 361 KEPAPTTPKEPAPTITTKPEPAPTITTKSAPTTPKEPAPTITTKKAPAPTTPKEPAPTTPKEPTP 420
QY 371 TTPKEPAPTTPKEPAPTAPKAPAPTTPKEPAPTITTKPEPAPTITTKESPTTPKEPAPTTPKE 430
Db 421 TTPKEPAPTTPKEPAPTAPKAPAPTTPKEPAPTITTKPEPAPTITTKESPTTPKEPAPTTPKE 480
QY 431 PAPTITTKSAPTITTKPEPAPTITTKSAPTTPKEPAPTITTKPEPAPTITTKKAPAPTTP 490

```



```
QY 570 PEEPAATPKAAAPNTPKAPATTPKEPAATTPKEPAATTPKETAATTPKGTATPILKEP 629
D 661 PEEPAATPKAAAPNTPKAPATTPKEPAATTPKEPAATTPKETAATTPKGTATPILKEP 720
QY 630 APPTPKKPAKELAPITTTKEPTSTSDKPAATTPKGTATTPKEPAATTPKEPAATTPKG 689
D 721 APPTPKKPAKELAPITTTKEPTSTSDKPAATTPKGTATTPKEPAATTPKEPAATTPKG 780
QY 690 TAPTLKEPAATTPKAPKELAPITTTKEPTSTSDKPAATTPKETAATTPKETAATTPKEPAATTPK 749
D 781 TAPTLKEPAATTPKAPKELAPITTTKEPTSTSDKPAATTPKETAATTPKETAATTPKEPAATTPK 840
QY 750 KPAATPETPPTTSEVSTPTTKEPTTIHKSPDESPKSAEPTPKALENSPKPGVPT 809
D 841 KPAATPETPPTTSEVSTPTTKEPTTIHKSPDESPKSAEPTPKALENSPKPGVPT 900
QY 810 TKTPAATKPMWTTAKDKTTERDLRTPEITTAAPKMTKETATTTETTESKITATTTQV 869
D 901 TKTPAATKPMWTTAKDKTTERDLRTPEITTAAPKMTKETATTTETTESKITATTTQV 960
QY 870 TSITTDTPPKITTLTKITTLAPKVTITTKITTIITTEIMNKPDEETAKPKDRATNSKATTPK 929
D 961 TSITTDTPPKITTLTKITTLAPKVTITTKITTIITTEIMNKPDEETAKPKDRATNSKATTPK 1020
QY 930 POKPTKAPKPTSTKPKTWPVRVKPKTTPTRKMTSTMPBLNPTSRIAEAMLOTTTRPN 989
D 1021 POKPTKAPKPTSTKPKTWPVRVKPKTTPTRKMTSTMPBLNPTSRIAEAMLOTTTRPN 1080
QY 990 QTPNSKLVEVPKSEDAGGAGETPHMLLRPHVFMPEVTPDMDYLRVPVNOGIIINPMLS 1049
D 1081 QTPNSKLVEVPKSEDAGGAGETPHMLLRPHVFMPEVTPDMDYLRVPVNOGIIINPMLS 1140

RESULT 6
US-07-757-022B-2
; Sequence 2, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/07757,022B
; APPLICATION NUMBER: US/07757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
```

```
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-2

Query Match 99.5%; Score 5546.9; DB 4; Length 1404;
Best Local Similarity 92.0%; Pred. No. 4.1e-169;
Matches 1049; Conservative 0; Mismatches 0; Indels 91; Gaps 2;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25
D 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCNDYNQHYMECCPDF 60
QY 26 -----ELCKGRCFESFERGRCDCDAQCKKYDKCCPDYESFCA----- 64
D 61 KRVTAEISCKGRCFESFERGRCDCDAQCKKYDKCCPDYESFCAEYHNFTSPSSKKAP 120
QY 65 -----EEHSVSENOE-----EEHSVSENOE-----EEHSVSENOE----- 89
D 121 PPGASGATIKSTTKRSKPPNKKKTKVIESEBITEEHSVSENOE-----EEHSVSENOE----- 180
QY 90 KIKSSNSAANRELQKKLVKDNKKNRTKKKTPKPPVDEAGSLDNGDFKVTTPDTST 149
D 181 KIKSSNSAANRELQKKLVKDNKKNRTKKKTPKPPVDEAGSLDNGDFKVTTPDTST 240
QY 150 TQHNKYSTPKITTTAKPINRPSLPNDSKTSKETSIVNKEVETVETTTNKQSTDG 209
D 241 TQHNKYSTPKITTTAKPINRPSLPNDSKTSKETSIVNKEVETVETTTNKQSTDG 300
QY 210 KEKTTSAKETOSIEKTSKOLAPTSKVLAKTPKAETTKGPAULTTKEPTTTPKBPAS 269
D 301 KEKTTSAKETOSIEKTSKOLAPTSKVLAKTPKAETTKGPAULTTKEPTTTPKBPAS 360
QY 270 TTPKEPTPTTIKSAPTTTPKEPAATTTKSAPTTPKEPAATTTTPKEPAATTTTPKEP 329
D 361 TTPKEPTPTTIKSAPTTTPKEPAATTTKSAPTTPKEPAATTTTPKEPAATTTTPKEP 420
QY 330 APPTTKSAPTTTPKEPAATTTKSAPTTPKEPAATTTTPKEPAATTTTPKEPAATTPK 389
D 421 APPTTKSAPTTTPKEPAATTTKSAPTTPKEPAATTTTPKEPAATTTTPKEPAATTPK 480
QY 390 EPAPTAPKPAATTPKEPAATTTTPKEPAATTTTPKEPAATTTTPKEPAATTTTPKEP 449
D 481 EPAPTAPKPAATTPKEPAATTTTPKEPAATTTTPKEPAATTTTPKEPAATTTTPKEP 540
QY 450 TTKSAPTTTPKEPSPTTTPKEPAATTTTPKPAATTTTPKPAATTTTPKEPAATTTTPK 509
D 541 TTKSAPTTTPKEPSPTTTPKEPAATTTTPKPAATTTTPKPAATTTTPKEPAATTTTPK 600
QY 510 APAPTAPKPAATTPKETAATTPPKLTPPTPEKLABTTPEKAPATTPPELAATTPPEPT 569
D 601 APAPTAPKPAATTPKETAATTPPKLTPPTPEKLABTTPEKAPATTPPELAATTPPEPT 660
QY 570 PEEPAATPKAAAPNTPKAPATTPKEPAATTPKEPAATTPKETAATTPKGTATPILKEP 629
D 661 PEEPAATPKAAAPNTPKAPATTPKEPAATTPKEPAATTPKETAATTPKGTATPILKEP 720
QY 630 APPTPKKPAKELAPITTTKEPTSTSDKPAATTPKGTATTPKEPAATTPKEPAATTPKG 689
D 721 APPTPKKPAKELAPITTTKEPTSTSDKPAATTPKGTATTPKEPAATTPKEPAATTPKG 780
QY 690 TAPTLKEPAATTPKAPKELAPITTTKEPTSTSDKPAATTPKETAATTPKETAATTPKEPAATTPK 749
D 781 TAPTLKEPAATTPKAPKELAPITTTKEPTSTSDKPAATTPKETAATTPKETAATTPKEPAATTPK 840
```



QY 750 KPAPTTPETPPPTTSEVSTPTTTKEPTTTHKSPDESTPELSABPTPKALENSPKBPGVPT 809  
DB 841 KPAPTTPETPPPTTSEVSTPTTTKEPTTTHKSPDESTPELSABPTPKALENSPKBPGVPT 900  
QY 810 TKTPAAATKPEMTTAKDKTTERDLRTPTPETTTAAPKMKTKETATTTKTTESKITATTQV 869  
DB 901 TKTPAAATKPEMTTAKDKTTERDLRTPTPETTTAAPKMKTKETATTTKTTESKITATTQV 960  
QY 870 TSTTTQDTPPKITTLKTTTLAPKVITTKTITTEIMNKPEBETAKPKDRATNSKATTPK 929  
DB 961 TSTTTQDTPPKITTLKTTTLAPKVITTKTITTEIMNKPEBETAKPKDRATNSKATTPK 1020  
QY 930 POKPTKAPKPTSTGKPKMVRKPKTPTPRKWTSTMBELNPTSRIAEAMLQTTTRPN 989  
DB 1021 POKPTKAPKPTSTGKPKMVRKPKTPTPRKWTSTMBELNPTSRIAEAMLQTTTRPN 1080  
QY 990 QTPNSKLVEVNPKSEDAAGAEGETPHMLRPHVFMPEVTPDMDYLPRVNPQGIIINPMLS 1049  
DB 1081 QTPNSKLVEVNPKSEDAAGAEGETPHMLRPHVFMPEVTPDMDYLPRVNPQGIIINPMLS 1140

RESULT 7

US-07-757-022B-62  
/ Sequence 62, Application US/07757022B  
/ Patent No. 6433142  
/ GENERAL INFORMATION:  
/ APPLICANT: Gesner, Thomas G.  
/ APPLICANT: Clark, Stephen C.  
/ APPLICANT: Turner, Katherine  
/ APPLICANT: Hewick, Rodney M.  
/ TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
/ NUMBER OF SEQUENCES: 143  
/ CORRESPONDENCE ADDRESS: 143  
/ ADDRESSEE: Genetics Institute, Inc.  
/ STREET: 87 CambridgePark Drive  
/ CITY: Cambridge  
/ STATE: Massachusetts  
/ COUNTRY: U.S.A.  
/ ZIP: 02140  
/ COMPUTER READABLE FORM:  
/ MEDIUM TYPE: Floppy disk  
/ COMPUTER: IBM PC compatible  
/ OPERATING SYSTEM: PC-DOS/MS-DOS  
/ SOFTWARE: PatentIn Release #1.0, Version #1.25  
/ CURRENT APPLICATION DATA:  
/ APPLICATION NUMBER: US/07/757,022B  
/ FILING DATE: 19910910  
/ CLASSIFICATION: 530  
/ PRIOR APPLICATION DATA:  
/ APPLICATION NUMBER: US 07/643,502  
/ FILING DATE: 18-JAN-1991  
/ PRIOR APPLICATION DATA:  
/ APPLICATION NUMBER: US 07/546,114  
/ FILING DATE: 29-JUN-1990  
/ PRIOR APPLICATION DATA:  
/ APPLICATION NUMBER: US 07/457,196  
/ FILING DATE: 29-DEC-1989  
/ PRIOR APPLICATION DATA:  
/ APPLICATION NUMBER: US 07/390,901  
/ FILING DATE: 08-AUG-1989  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: Cserr, Luann  
/ REGISTRATION NUMBER: 31,822  
/ REFERENCE/DOCKET NUMBER: GI 5190  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: (617)876-1170  
/ TELEFAX: (617)876-5851  
/ INFORMATION FOR SEQ ID NO: 62:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 1404 amino acids  
/ TYPE: AMINO ACID  
/ TOPOLOGY: linear

MOLECULE TYPE: protein  
US-07-757-022B-62

Query Match 99.5%; Score 5546.9; DB 4; Length 1404;  
Best Local Similarity 92.0%; Pred. No. 4.1e-169;  
Matches 1049; Conservative 0; Mismatches 0; Indels 91; Gaps 2;  
QY 1 MAWKTLPIYLILLLSVRFVIOQVSSQ-----25  
DB 1 MAWKTLPIYLILLLSVRFVIOQVSSQDLSSCAGRCGEGYSRDATCNCDYNQHYMECCPDF 60  
QY 26 -----ELSCGRCFCPSFERGRECDCAOQCKYDKCCPDYSEFCA-----64  
DB 61 KRVTAEISLCKGRCFCPSFERGRECDCAOQCKYDKCCPDYSEFCAEVHNTPSPSSKAP 120  
QY 65 -----BEHSVSENQESSSSSSSSSSSIW 89  
DB 121 PPSGASQITKSTTKRGPKNKKTKKVIESEBITTEHSVSENQESSSSSSSSSSSIW 180  
QY 90 KIKSSKNSAANRELOKKLVKONKKNRTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDTST 149  
DB 181 KIKSSKNSAANRELOKKLVKONKKNRTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDTST 240  
QY 150 TQHNKVSTSPKITTAKPINPRPSLPNSDTSKETSITVNKETTIVETKETTITNKQSTSDG 209  
DB 241 TQHNKVSTSPKITTAKPINPRPSLPNSDTSKETSITVNKETTIVETKETTITNKQSTSDG 300  
QY 210 KENTSAKETQSIEKTSKDLAPTSKVLAKEPTPKAETTTKGPALTTPKBPTTTPKEPAS 269  
DB 301 KENTSAKETQSIEKTSKDLAPTSKVLAKEPTPKAETTTKGPALTTPKBPTTTPKEPAS 360  
QY 270 TTPKEPTPTTIKSAPTTKEPATTTTKSAPTTKEPATTTKEPATTTKEPATTTKEPATTTKEP 329  
DB 361 TTPKEPTPTTIKSAPTTKEPATTTTKSAPTTKEPATTTKEPATTTKEPATTTKEPATTTKEP 420  
QY 330 APITTKSAPTTTPKEPATTTPKKPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPK 389  
DB 421 APITTKSAPTTTPKEPATTTPKKPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPK 480  
QY 390 EPAPTAPKKPAPTTTPKEPATTTTPKEPATTTTKBPSPTTPKEPATTTTKSAPTTTKBAPT 449  
DB 481 EPAPTAPKKPAPTTTPKEPATTTTPKEPATTTTKBPSPTTPKEPATTTTKSAPTTTKBAPT 540  
QY 450 TTKSAPTTTPKEPSPTTTKEPATTTTPKEPATTTPKKPATTTTPKEPATTTTPKEPATTTTKP 509  
DB 541 TTKSAPTTTPKEPSPTTTKEPATTTTPKEPATTTPKKPATTTTPKEPATTTTPKEPATTTTKP 600  
QY 510 APTAPKEPATTTPKETAPTTPKKLTPTTPEKLAPTTTPKEPATTTTPKEPATTTPEEPTPTT 569  
DB 601 APTAPKEPATTTPKETAPTTPKKLTPTTPEKLAPTTTPKEPATTTTPKEPATTTPEEPTPTT 660  
QY 570 PEEPAPTTPKAAAPNTPKEPATTTTPKEPATTTTPKEPATTTTPKETAPTTPKGTAPTTLKBP 629  
DB 661 PEEPAPTTPKAAAPNTPKEPATTTTPKEPATTTTPKEPATTTTPKETAPTTPKGTAPTTLKBP 720  
QY 630 APTTPKKPAPKELAPTTTKEPTSTSDKAPTTPKGTAPTTPKEPATTTTPKEPATTTPKG 689  
DB 721 APTTPKKPAPKELAPTTTKEPTSTSDKAPTTPKGTAPTTPKEPATTTTPKEPATTTPKG 780  
QY 690 TAPTTLKEPATTTTPKKPAPKELAPTTTKGPTSTSDKAPTTTPKETAPTTPKEPATTTTPK 749  
DB 781 TAPTTLKEPATTTTPKKPAPKELAPTTTKGPTSTSDKAPTTTPKETAPTTPKEPATTTTPK 840  
QY 750 KPAPTTPETPPPTTSEVSTPTTTKEPTTTHKSPDESTPELSABPTPKALENSPKBPGVPT 809  
DB 841 KPAPTTPETPPPTTSEVSTPTTTKEPTTTHKSPDESTPELSABPTPKALENSPKBPGVPT 900  
QY 810 TKTPAAATKPEMTTAKDKTTERDLRTPTPETTTAAPKMKTKETATTTKTTESKITATTQV 869  
DB 901 TKTPAAATKPEMTTAKDKTTERDLRTPTPETTTAAPKMKTKETATTTKTTESKITATTQV 960  
QY 870 TSTTTQDTPPKITTLKTTTLAPKVITTKTITTEIMNKPEBETAKPKDRATNSKATTPK 929

Db 961 TSTTTQDTTTPFKITTLTKTTTTLAPKVTTTKKTIITTEIMNKPEETAKPKDRATNSKATTPK 1020  
Qy 930 POKPTKAPKPTSTKKPKMTMPVRKPKTTPTPKMTSTMPBLNPTSRIAEAMLOTTTRPN 989  
Db 1021 POKPTKAPKPTSTKKPKMTMPVRKPKTTPTPKMTSTMPBLNPTSRIAEAMLOTTTRPN 1080  
Qy 990 QTNSKLIVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMOYLPRVFNQGGIINPMLS 1049  
Db 1081 QTNSKLIVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMOYLPRVFNQGGIINPMLS 1140

RESULT 8  
US-09-298-970A-1  
; Sequence 1, Application US/092989970A  
; Patent No. 6743774  
; GENERAL INFORMATION:  
; APPLICANT: Jay, Gregory D.  
; TITLE OF INVENTION: TRIBONECTINS  
; FILE REFERENCE: 21486-026  
; CURRENT APPLICATION NUMBER: US/09/298,970A  
; CURRENT FILING DATE: 2001-06-19  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: Fast-Seq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 1404  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-298-970A-1

Query Match 99.5%; Score 5546.9; DB 4; Length 1404;  
Best Local Similarity 92.0%; Pred. No. 4.1e-169;  
Matches 1049; Conservative 0; Mismatches 0; Indels 91; Gaps 2;  
Qy 1 MAWKTLPIYLLLSLVFVIQQVSSQ----- 25  
Db 1 MAWKTLPIYLLLSLVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCNDYNCQHYMECCPDF 60  
Qy 26 -----ELSCCKGRCFESFERGREGCDCAOCKKYDKCCPDYESFCA----- 64  
Db 61 KRVTAEIASCGRCFESFERGREGCDCAOCKKYDKCCPDYESFCAEVHNFTSPSSKAP 120  
Qy 65 -----BEHSVSNQESSSSSSSSSSSTIW 89  
Db 121 PPSGASQTIKSTTKRSPKPNKKTKKVBSEBITEHSVSNQESSSSSSSSSSSSSTIW 180

Qy 90 KIKSSKNSAANRELQKKLVKDNKNKRTKKPKPPVVDVDEAGSLDNGDFKVTTPDTST 149  
Db 181 KIKSSKNSAANRELQKKLVKDNKNKRTKKPKPPVVDVDEAGSLDNGDFKVTTPDTST 240  
Qy 150 TQHNKYSTSPKITTAKEINPRPSLPNNSDTSKETSITVNETTVEKETTITNKQSTDG 209  
Db 241 TQHNKYSTSPKITTAKEINPRPSLPNNSDTSKETSITVNETTVEKETTITNKQSTDG 300  
Qy 210 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTKAEATTTKGPALTTTPKEPTTPPKBPAS 269  
Db 301 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTKAEATTTKGPALTTTPKEPTTPPKBPAS 360  
Qy 270 TTPKEPTPTTIKSAPTTPKBPATTTKSAPTTKBPAPTTTKBPAPTTTKBPAPTTTKBP 329  
Db 361 TTPKEPTPTTIKSAPTTPKBPAPTTTKSAPTTPKBPAPTTTKBPAPTTTKBPAPTTTKBP 420  
Qy 330 APITTTKSAPTTPKBPAPTTPKKAPTTPKBPAPTTPKPEPTTPPKBPAPTTPKBPAPTTPK 389  
Db 421 APITTTKSAPTTPKBPAPTTPKKAPTTPKBPAPTTPKPEPTTPPKBPAPTTPKBPAPTTPK 480  
Qy 390 EPAPAPKAPAPTTPKBPAPTTPKBPAPTTTKBSPPTTKBPAPTTTKSAPTTTKBPAPTT 449  
Db 481 EPAPAPKAPAPTTPKBPAPTTPKBPAPTTTKBSPPTTKBPAPTTTKSAPTTTKBPAPTT 540  
Qy 450 TTKSAPTTPKBSPPTTKBPAPTTPKBPAPTTPKKAPAPTTPKBPAPTTPKBPAPTTTKKP 509  
Db 541 TTKSAPTTPKBSPPTTKBPAPTTPKBPAPTTPKKAPAPTTPKBPAPTTPKBPAPTTTKKP 600

Qy 510 APTAPKEPAPTTPKETAPTTPKKLTPPTPEKLAPTTPKEKAPATTPPEELAPTTPEEPPTPT 569  
Db 601 APTAPKEPAPTTPKETAPTTPKKLTPPTPEKLAPTTPKEKAPATTPPEELAPTTPEEPPTPT 660  
Qy 570 PEPAPPTPKAAAPNTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKGTAPTTTILKEP 629  
Db 661 PEPAPPTPKAAAPNTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKGTAPTTTILKEP 720  
Qy 630 APITTPKAPKAPKELAPTTTKETPTSTSDKAPATTPKGTAPTTPKBPAPTTPKBPAPTTPKG 689  
Db 721 APITTPKAPKAPKELAPTTTKETPTSTSDKAPATTPKGTAPTTPKBPAPTTPKBPAPTTPKG 780  
Qy 690 TAPTTILKEPAPTTPKKAPKELAPTTTKGPTSTTSKAPATTPPKETAPTTPKBPAPTTPK 749  
Db 781 TAPTTILKEPAPTTPKKAPKELAPTTTKGPTSTTSKAPATTPPKETAPTTPKBPAPTTPK 840  
Qy 750 KPAPTTPEPTPTTSEVSTPTTTKBPPTTIHKSDESTPELSASPTPKALNSPKBPVPT 809  
Db 841 KPAPTTPEPTPTTSEVSTPTTTKBPPTTIHKSDESTPELSASPTPKALNSPKBPVPT 900  
Qy 810 TKTPAATKPEMTTAKDKTERDLRTTPETTTAAPKMTKETATTTTEKTTESKITATTTQV 869  
Db 901 TKTPAATKPEMTTAKDKTERDLRTTPETTTAAPKMTKETATTTTEKTTESKITATTTQV 960  
Qy 870 TSTTTQDTTTPFKITTLTKTTTTLAPKVTTTKKTIITTEIMNKPEETAKPKDRATNSKATTPK 929  
Db 961 TSTTTQDTTTPFKITTLTKTTTTLAPKVTTTKKTIITTEIMNKPEETAKPKDRATNSKATTPK 1020  
Qy 930 POKPTKAPKPTSTKKPKMTMPVRKPKTTPTPKMTSTMPBLNPTSRIAEAMLOTTTRPN 989  
Db 1021 POKPTKAPKPTSTKKPKMTMPVRKPKTTPTPKMTSTMPBLNPTSRIAEAMLOTTTRPN 1080  
Qy 990 QTNSKLIVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMOYLPRVFNQGGIINPMLS 1049  
Db 1081 QTNSKLIVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMOYLPRVFNQGGIINPMLS 1140

## RESULT 9

US-10-164-595-78  
; Sequence 78, Application US/10164595  
; Patent No. 6657054  
; GENERAL INFORMATION:  
; APPLICANT: Origene Technologies, Inc  
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides  
; FILE REFERENCE: IU 103 R1  
; CURRENT APPLICATION NUMBER: US/10/164,595  
; CURRENT FILING DATE: 2002-06-10  
; NUMBER OF SEQ ID NOS: 80  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 78  
; LENGTH: 1404  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-164-595-78

Query Match 99.1%; Score 5523.9; DB 4; Length 1404;  
Best Local Similarity 91.8%; Pred. No. 2.2e-168;  
Matches 1045; Conservative 0; Mismatches 3; Indels 91; Gaps 2;  
Qy 1 MAWKTLPIYLLLSLVFVIQQVSSQ----- 25  
Db 1 MAWKTLPIYLLLSLVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCNDYNCQHYMECCPDF 60  
Qy 26 -----ELSCCKGRCFESFERGREGCDCAOCKKYDKCCPDYESFCA----- 64  
Db 61 KRVTAEIASCGRCFESFERGREGCDCAOCKKYDKCCPDYESFCAEVHNFTSPSSKAP 120  
Qy 65 -----BEHSVSNQESSSSSSSSSSSTIW 89  
Db 121 PPSGASQTIKSTTKRSPKPNKKTKKVBSEBITEHSVSNQESSSSSSSSSSSTIW 180  
Qy 90 KIKSSKNSAANRELQKKLVKDNKNKRTKKPKPPVVDVDEAGSLDNGDFKVTTPDTST 149

181 KIKSSKNSAANRELQKKLVKDNKNRTKKKTPKPPVVDAGSLDNGDFKVTTPDTST 240  
150 TQHNKUSTSKITITAKINPRPSLPNSDTSKETSUWNVKETTVEKETTITNKQISTDG 209  
241 TQHNKUSTSKITITAKINPRPSLPNSDTSKETSUWNVKETTVEKETTITNKQISTDG 300  
210 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTTPKEPAS 269  
301 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTTPKEPAS 360  
270 TTPKEPTPTTIKSAPTTPKBPAPTTTSGAPTTKBPAPTTTTPKEPTTTPKEPAS 329  
361 TTPKEPTPTTIKSAPTTPKBPAPTTTSGAPTTKBPAPTTTTPKEPTTTPKEPAS 420  
330 APPTTKSAPTTKBPAPTTTPKBPAPTTTPKEPTTTPKEPTTTPKEPTTTPKEPTTTPK 389  
421 APPTTKSAPTTKBPAPTTTPKBPAPTTTPKEPTTTPKEPTTTPKEPTTTPKEPTTTPK 480  
390 EPAPTAPKBPAPTTTPKEPTTTPKEPTTTPKEPTTTPKEPTTTPKEPTTTPKEPTTTPK 449  
481 EPAPTAPKBPAPTTTPKEPTTTPKEPTTTPKEPTTTPKEPTTTPKEPTTTPKEPTTTPK 540  
450 TTKSAPTTTPKEPTTTPKEPTTTPKEPTTTPKEPTTTPKEPTTTPKEPTTTPKEPTTTPK 509  
541 TTKSAPTTTPKEPTTTPKEPTTTPKEPTTTPKEPTTTPKEPTTTPKEPTTTPKEPTTTPK 600  
510 APAPTAPKBPAPTTTPKEPTTTPKEPTTTPKEPTTTPKEPTTTPKEPTTTPKEPTTTPK 569  
601 APAPTAPKBPAPTTTPKEPTTTPKEPTTTPKEPTTTPKEPTTTPKEPTTTPKEPTTTPK 660  
570 PEPAPTTPKAAAPNTPKBPAPTTTPKEPTTTPKEPTTTPKEPTTTPKEPTTTPKEPTTTPK 629  
661 PEPAPTTPKAAAPNTPKBPAPTTTPKEPTTTPKEPTTTPKEPTTTPKEPTTTPKEPTTTPK 720  
630 APPTPKAPKELAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTPKG 689  
721 APPTPKAPKELAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTPKG 780  
690 TAPTTLKEPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPK 749  
781 TAPTTLKEPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPK 840  
750 KPAPTTPETPTTSEVSTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTT 809  
841 KPAPTTPETPTTSEVSTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTT 900  
810 TKTPAAATKPEMTTAKDKTTERDLRTTPETTTAAPKMTKETAITTEKTTESKITATTQV 869  
901 TKTPAAATKPEMTTAKDKTTERDLRTTPETTTAAPKMTKETAITTEKTTESKITATTQV 960  
870 TSTTTQDTPPKITLTKITLAPKVTITTKITITTEIMNKEETAKPKDRATNSKATTPK 929  
961 TSTTTQDTPPKITLTKITLAPKVTITTKITITTEIMNKEETAKPKDRATNSKATTPK 1020  
930 PQKPTKAPKPTSTTKPKPMVRVKPXTTTPPKMTSTMPELNPTSTIAEAMQTTTRPN 989  
1021 PQKPTKAPKPTSTTKPKPMVRVKPXTTTPPKMTSTMPELNPTSTIAEAMQTTTRPN 1080  
990 QTNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMYLRPVNPQGIINPMLS 1049  
1081 QTNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMYLRPVNPQGIINPMLS 1140

RESULT 10

US-07-757-022B-50  
; Sequence 50, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 CambridgePark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/757,022B  
FILING DATE: 19910910  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Cseri, Luann  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 5190  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)876-1170  
TELEFAX: (617)876-5851  
INFORMATION FOR SEQ ID NO: 50:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1314 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-757-022B-50  
Query Match 97.7%; Score 5447.9; DB 4; Length 1314;  
Best Local Similarity 97.8%; Pred. No. 5.4e-166;  
Matches 1027; Conservative 7; Mismatches 15; Indels 1; Gaps 1;  
QY 1 MANKTLPYLLLLSVFVIOQVSSQEL-SCKGRCFESFERGRCDCDAQCKYDKCCPDY 59  
DB 1 MANKTLPYLLLLSVFVIOQVSSQELSSCAGCGEGYSDATCNCYNCQHWECPCDF 60  
QY 60 ESFCAEHSVSNEQSSSSSSSSSSSSSTIWKIKSSKNSAANRELQKKLVKDNKNRTKK 119  
DB 61 KRVCYAEHSVSNEQSSSSSSSSSSSSSTIWKIKSSKNSAANRELQKKLVKDNKNRTKK 120  
QY 120 KPTPKPPVDEAGSLDNGDFKVTTPDTSTTQHNKUSTSKITITAKINPRPSLPNSDT 179  
DB 121 KPTPKPPVDEAGSLDNGDFKVTTPDTSTTQHNKUSTSKITITAKINPRPSLPNSDT 180  
QY 180 SKTSTLVNKETTVEKETTITNKQISTDGKEKTTSAKETQSIEKTSKADLAPTSKVLAK 239  
DB 181 SKTSTLVNKETTVEKETTITNKQISTDGKEKTTSAKETQSIEKTSKADLAPTSKVLAK 240  
QY 240 PTPKAETTTKGPALTTKPEPTTTPKEPTTTPKEPTTTPKEPTTTPKEPTTTPKEPTTTPK 299  
DB 241 PTPKAETTTKGPALTTKPEPTTTPKEPTTTPKEPTTTPKEPTTTPKEPTTTPKEPTTTPK 300  
QY 300 TTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKE 359  
DB 301 TTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKE 360  
QY 360 PAPTTPKEPTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKE 419

```
Db 361 PATTTPKEPTTTPKEPAPTTTKEPAPTTTKEPAPTTAPKKPAPTTTKEPAPTTTKEPAPTT 420
Qy 420 TKPSPTTTPKEPAPTTTTSAPTTTKEPAPTTTTSAPTTTKEPAPTTTKEPAPTTTKEPAP 479
Db 421 TKPSPTTTPKEPAPTTTTSAPTTTKEPAPTTTTSAPTTTKEPAPTTTKEPAPTTTKEPAP 480
Qy 480 TTPKKPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 539
Db 481 TTPKKPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 540
Qy 540 KLAPTTPEKAPTTTPEELAPTTTPEEPPTTPEEPAPTTTAKAAPTNTKEPAPTTTKEPAP 599
Db 541 KLAPTTPEKAPTTTPEELAPTTTPEEPPTTPEEPAPTTTAKAAPTNTKEPAPTTTKEPAP 600
Qy 600 TTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 659
Db 601 TTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 660
Qy 660 PTPKGTAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 719
Db 661 PTPKGTAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 720
Qy 720 TSTTSKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 779
Db 721 TSTTSKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 780
Qy 780 KSDSTPELSAPTTKALNSPKPGVPTTKPAATKPEMTTAKDKTTTERDLRTTPT 839
Db 781 KSDSTPELSAPTTKALNSPKPGVPTTKPAATKPEMTTAKDKTTTERDLRTTPT 840
Qy 840 TTAAPKMTKATTTTETKTESKITATTTQVSTTTQDTTTPKTIITLAKTTTLAPKVTTTK 899
Db 841 TTAAPKMTKATTTTETKTESKITATTTQVSTTTQDTTTPKTIITLAKTTTLAPKVTTTK 900
Qy 900 TITTEIMNKPEETAKPKDRATNSKATTPPKPQKPTKAPKPTSTSKPKTTPVRVKPKTTP 959
Db 901 TITTEIMNKPEETAKPKDRATNSKATTPPKPQKPTKAPKPTSTSKPKTTPVRVKPKTTP 960
Qy 960 TPKMTSTMPELNPTSIAEAMLTQTTTPNQNTNSKLVNPNKSDAGGAEGETPHMLLR 1019
Db 961 TPKMTSTMPELNPTSIAEAMLTQTTTPNQNTNSKLVNPNKSDAGGAEGETPHMLLR 1020
Qy 1020 PHVFMPEVTTPDMYLPVRPNQGIINPMLS 1049
Db 1021 PHVFMPEVTTPDMYLPVRPNQGIINPMLS 1050
```

## RESULT 11

```
US-07-757-022B-46
; Sequence 46, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07757,022B
```

```
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: G1 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1320 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-46
```

```
Query Match 96.7%; Score 5393.3; DB 4; Length 1320;
Best Local Similarity 96.1%; Pred. No. 3e-164;
Matches 1015; Conservative 8; Mismatches 26; Indels 7; Gaps 1;

Qy 1 MAWKTLPIYLLLSLVFVIQVSSQELSCKGRCEFSERGERBCDCAQCKYKCCPDYE 60
Db 1 MAWKTLPIYLLLSLVFVIQVSSQELSCKGRCEFSERGERBCDCAQCKYKCCPDYE 60
Qy 61 SFCAEHSVSENGESSSSSSSTIWKIKSKNSAANKLQKLU-----KVKNK 113
Db 61 SFCAEVNPTSPSSSKKAPPPSGASQTIKSTTKRSPKPNKKTKKVIKSEIEVKNK 120
Qy 114 KNRTKKKPTPKPPVVDGAGSLDNGDFKVTTPDTISTTQHNKVNSTSPKITTAKPINPRESL 173
Db 121 KNRTKKKPTPKPPVVDGAGSLDNGDFKVTTPDTISTTQHNKVNSTSPKITTAKPINPRESL 180
Qy 174 PPNSDTSKETSLTNKETTIVETKETTITNNKQSTTDGKEKTSKETSIEKTSKADLAPT 233
Db 181 PPNSDTSKETSLTNKETTIVETKETTITNNKQSTTDGKEKTSKETSIEKTSKADLAPT 240
Qy 234 SKVLAKPTPKAETTTKGPALTTKPEPTTTPKPEASTTPKPEPTTTTIKSAPTTPKBPAPT 293
Db 241 SKVLAKPTPKAETTTKGPALTTKPEPTTTPKPEASTTPKPEPTTTTIKSAPTTKEBPAPT 300
Qy 294 TTKSAPTTPKBPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 353
Db 301 TTKSAPTTPKBPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 360
Qy 354 PTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 413
Db 361 PTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 420
Qy 414 EPAPTTTKEPSPPTTKEPAPTTTTSAPTTTKEPAPTTTTSAPTTTKEPAPTTTKEPAPTTTKE 473
Db 421 EPAPTTTKEPSPPTTKEPAPTTTTSAPTTTKEPAPTTTTSAPTTTKEPAPTTTKEPAPTTTKE 480
Qy 474 PKBPAPTTPKBPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 533
Db 481 PKBPAPTTPKBPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 540
Qy 534 TPTTPEKLAPTTPEKAPTTTPEELAPTTTPEEPPTTPEEPAPTTTAKAAPTNTKEPAPTTTKE 593
Db 541 TPTTPEKLAPTTPEKAPTTTPEELAPTTTPEEPPTTPEEPAPTTTAKAAPTNTKEPAPTTTKE 600
```



```
QY 834 RTTPEITTAAPKMTKETATTTTEKTESKITATTTQVSTTTQDTPPKITLTKTTTLAPK 893
Db 841 RTTPEITTAAPKMTKETATTTTEKTESKITATTTQVSTTTQDTPPKITLTKTTTLAPK 900
QY 894 VTTTTKITITTEIMNKPEETAAPKDRATNSKATTPKQKPTKAPKXSTSKPKTMRVR 953
Db 901 VTTTTKITITTEIMNKPEETAAPKDRATNSKATTPKQKPTKAPKXSTSKPKTMRVR 960
QY 954 KPXTTTPRKWSTMPNLNPTSRIAEAMLOTTTRPNQTPNSKLVEVNPKSEDAGAGSET 1013
Db 961 KPXTTTPRKWSTMPNLNPTSRIAEAMLOTTTRPNQTPNSKLVEVNPKSEDAGAGSET 1020
QY 1014 PHMLLRPHVFMPEVTPDMVYLPRVFNQGIINPMLS 1049
Db 1021 PHMLLRPHVFMPEVTPDMVYLPRVFNQGIINPMLS 1056

RESULT 13
US-10-164-595-58
; Sequence 58, Application US/10164595
; Patent No. 6657054
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
; FILE REFERENCE: IU 103 R1
; CURRENT APPLICATION NUMBER: US/10/164,595
; CURRENT FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 58
; LENGTH: 1320
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-164-595-58

Query Match 96.6%; Score 5384.3; DB 4; Length 1320;
Best Local Similarity 95.9%; Pred. No. 5.8e-164;
Matches 1013; Conservative 8; Mismatches 28; Indels 7; Gaps 1;

QY 1 MAWKTLPIYLLLSVFIQVVSQELSCKGRCFESFERGRCDCDAQCKYDKCCPDYE 60
Db 1 MAWKTLPIYLLLSVFIQVVSQELSCKGRCFESFERGRCDCDAQCKYDKCCPDYE 60
QY 61 SFCAEHSVSENQESSSSSSSSSTTWIKSSKNSANRELQKL-----KVQDNK 113
Db 61 SFCAEHSVSPSSKKAPPPSGASQTIKSTKRSKPPNKKTKKVISEITEVDKNK 120
QY 114 KNRTKKKPTPKPPVVDAGSLDNGDFKVTTPDTSSTQHNVSTSPKITTAKPINRPSL 173
Db 121 KNRTKKKPTPKPPVVDAGSLDNGDFKVTTPDTSSTQHNVSTSPKITTAKPINRPSL 180
QY 174 PPSDTSKESLTVNKETTVETKTTTNTKQTSDDGKEKTTSAKETQSIEKTSADLAPT 233
Db 181 PPSDTSKESLTVNKETTVETKTTTNTKQTSDDGKEKTTSAKETQSIEKTSADLAPT 240
QY 234 SKVLAKPTKAEITTKGPAITTPKEPTTPKEASTTPKEPTTTIKSAPTTPKEAPT 293
Db 241 SKVLAKPTKAEITTKGPAITTPKEPTTPKEASTTPKEPTTTIKSAPTTPKEAPT 300
QY 294 TTKSAPTTPKEAPTTPKPAITTPKEAPTTPKPAITTPKPAITTPKPAITTPKPA 353
Db 301 TTKSAPTTPKEAPTTPKPAITTPKEAPTTPKPAITTPKPAITTPKPAITTPKPA 360
QY 354 PTPKPAITTPKEAPTTPKPAITTPKPAITTPKPAITTPKPAITTPKPAITTPK 413
Db 361 PTPKPAITTPKEAPTTPKPAITTPKPAITTPKPAITTPKPAITTPKPAITTPK 420
QY 414 EPAPTTPKPSPTTPKPAITTPKPAITTPKPAITTPKPAITTPKPAITTPKPA 473
Db 421 EPAPTTPKPSPTTPKPAITTPKPAITTPKPAITTPKPAITTPKPAITTPKPA 480
QY 474 PKEAPTTPKPAITTPKPAITTPKPAITTPKPAITTPKPAITTPKPAITTPK 533
```

```
Db 481 PKEAPTTPKPAITTPKPAITTPKPAITTPKPAITTPKPAITTPKPAITTPK 540
QY 534 TPTTPEKLAAPTTEKPAITTPPELAPTTPEEPPTTPPEEPAPTTPKAAANTPKPA 593
Db 541 TPTTPEKLAAPTTEKPAITTPPELAPTTPEEPPTTPPEEPAPTTPKAAANTPKPA 600
QY 594 PKEAPTTPKPAITTPKPAITTPKPAITTPKPAITTPKPAITTPKPAITTPKPA 653
Db 601 PKEAPTTPKPAITTPKPAITTPKPAITTPKPAITTPKPAITTPKPAITTPKPA 660
QY 654 TSOKPAITTPKGAITTPKPAITTPKPAITTPKGAITTPKGAITTPKGAITTP 713
Db 661 TCDKPAITTPKGAITTPKPAITTPKPAITTPKGAITTPKGAITTPKGAITTP 720
QY 714 TTTKGPTSTSDKPAITTPKPAITTPKPAITTPKPAITTPKPAITTPKPAITTP 773
Db 721 TTTKGPTSTSDKPAITTPKPAITTPKPAITTPKPAITTPKPAITTPKPAITTP 780
QY 774 EPTTIHKSPDESTEPKALNSPKPEPGVPTTKTAAATKPEMTTTAKDKTTBRDL 833
Db 781 EPTTIHKSPDESTEPKALNSPKPEPGVPTTKTAAATKPEMTTTAKDKTTBRDL 840
QY 834 RTTPEITTAAPKMTKETATTTTEKTESKITATTTQVSTTTQDTPPKITLTKTT 893
Db 841 RTTPEITTAAPKMTKETATTTTEKTESKITATTTQVSTTTQDTPPKITLTKTT 900
QY 894 VTTTTKITITTEIMNKPEETAAPKDRATNSKATTPKQKPTKAPKXSTSKPKTMRVR 953
Db 901 VTTTTKITITTEIMNKPEETAAPKDRATNSKATTPKQKPTKAPKXSTSKPKTMRVR 960
QY 954 KPXTTTPRKWSTMPNLNPTSRIAEAMLOTTTRPNQTPNSKLVEVNPKSEDAGAGSET 1013
Db 961 KPXTTTPRKWSTMPNLNPTSRIAEAMLOTTTRPNQTPNSKLVEVNPKSEDAGAGSET 1020
QY 1014 PHMLLRPHVFMPEVTPDMVYLPRVFNQGIINPMLS 1049
Db 1021 PHMLLRPHVFMPEVTPDMVYLPRVFNQGIINPMLS 1056

RESULT 14
US-07-757-022B-40
; Sequence 40, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/07757,022B
; APPLICATION NUMBER: US/07757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1361 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-40

Query Match 36.5%; Score 5379.2; DB 4; Length 1361;
Best local Similarity 92.5%; Pred. No. 8.7e-164;
Matches 1015; Conservative 8; Mismatches 26; Indels 48; Gaps 2;

QY 1 MAWKTLPIYLLLSLVFVIOQVSSQ----- 25
DB 1 MAWKTLPIYLLLSLVFVIOQVSSQDLSSCAGRGEGYSDATCNDYCNQHMECCPDF 60

QY 26 -----ELSCGRCFESFERGREGDCDAQCKYDKCCPDYSEFCAEHSVSENQESSSS 79
DB 61 KRVCIAELSCGRCFESFERGREGDCDAQCKYDKCCPDYSEFCAEHSVNTSPSSKKAP 120

QY 80 SSSSSSTIWKIKSSKNSAANRELQKL-----KVONKQKRTKKKPTPKPPVDEAG 132
DB 121 PPSGASQTIKTRSPKPPKTKKVI SEEITEVKDNKQKRTKKKPTPKPPVDEAG 180

QY 133 SGLDNGDFKVTTPDSTTQNKVSTSPKITTAKPINRPSLPPNSDTSKETSIVNKETT 192
DB 181 SGLDNGDFKVTTPDSTTQNKVSTSPKITTAKPINRPSLPPNSDTSKETSIVNKETT 240

QY 193 VETKEITTNKQISTDCKEKTTSKAKTQSIKTSKADLAFTSKVLAKPTPKAETTTKGPA 252
DB 241 VETKEITTNKQISTDCKEKTTSKAKTQSIKTSKADLAFTSKVLAKPTPKAETTTKGPA 300

QY 253 LITPKETPTTPKEPASTTPKEPTPTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKE 312
DB 301 LATPKETPTTPKEPASTTPKEPTPTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKE 360

QY 313 PAPTTPKEAPTTTKPEAPTTTKSAPTTKPEAPTTTPKPEAPTTTPKPEAPTTTKEPTPTT 372
DB 361 PAPTTPKEAPTTTKPEAPTTTKSAPTTKPEAPTTTPKPEAPTTTPKPEAPTTTKEPTPTT 420

QY 373 PKPEAPTTKEPAPTTPKEAPTTAPKPEAPTTTPKPEAPTTTPKPEAPTTTKPEFPTTKEPA 432
DB 421 PKPEAPTTKEPAPTTPKPEAPTTAPKPEAPTTTPKPEAPTTTPKPEAPTTTKPEFPTTKEPA 480

QY 433 PTTTKSAPTTTKPEAPTTTKSAPTTTPKEPSGTTTKEPAPTTPKPEAPTTTPKPEAPTTTKE 492
DB 481 PTTTKSAPTTTKPEAPTTTKSAPTTTPKEPSGTTTKEPAPTTPKPEAPTTTPKPEAPTTTKE 540

QY 493 PAPTTPKEAPTTTKPEAPTTAPKPEAPTTTPKETAFTPKLTPTTPPEKLAPTTTPEKDAPT 552
DB 541 PAPTTPKEAPTTTKPEAPTTAPKPEAPTTTPKETAFTPKLTPTTPPEKLAPTTTPEKDAPT 600

QY 553 TPPELAPTTPEEPTPTTPEEPAPTTPKAAAPNTPKPEAPTTTPKPEAPTTTPKPEAPTTTKE 612
DB 601 TPPELAPTTPEEPTPTTPEEPAPTTPKAAAPNTPKPEAPTTTPKPEAPTTTPKPEAPTTTKE 660

QY 613 TAPTPKGTAPTTLKPEAPTTPKKPAKELAPTTTKEPTSTSDKAPTTTPKGTAPTTPK 672
DB 661 TAPTPKGTAPTTLKPEAPTTPKKPAKELAPTTTKEPTSTSDKAPTTTPKGTAPTTPK 720

```

## RESULT 15

US-07-757-022B-74

; Sequence 74, Application US/07/757022B

; Patent No. 6433142

; GENERAL INFORMATION:

; APPLICANT: Gesner, Thomas G.

; APPLICANT: Clark, Stephen C.

; APPLICANT: Turner, Katherine

; APPLICANT: Hewick, Rodney M.

; TITLE OF INVENTION: Megakaryocyte Stimulating Factors

; NUMBER OF SEQUENCES: 143

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genetics Institute, Inc.

; STREET: 87 CambridgePark Drive

; CITY: Cambridge

; STATE: Massachusetts

; COUNTRY: U.S.A.

; ZIP: 02140

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/757,022B

; FILING DATE: 19910910

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/643,502

; FILING DATE: 18-JAN-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/546,114

; FILING DATE: 29-JUN-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/457,196

; FILING DATE: 29-DEC-1989

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/390,901

; FILING DATE: 08-AUG-1989

; ATTORNEY/AGENT INFORMATION:

; NAME: Cserr, Luann

; REGISTRATION NUMBER: 31,822

; REFERENCE/DOCKET NUMBER: GI 5190

; TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)876-1170  
TELEFAX: (617)876-5851  
INFORMATION FOR SEQ ID NO: 74:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1038 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-757-022B-74

Query Match 96.1%; Score 5358.7; DB 4; Length 1038;  
Best Local Similarity 95.9%; Pred. No. 2.8e-163;  
Matches 1006; Conservative 0; Mismatches 0; Indels 43; Gaps 1;  
QY 1 MAWKTLPIYLLLLSVFVIQVSSQELSKGRCFESFERGECDCDAQCKYDKCCPDYE 60  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 1 MAWKTLPIYLLLLSVFVIQVSSQELSKGRCFESFERGECDCDAQCKYDKCCPDYE 60  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 61 SFCNEHVSSENQESSSSSSSSSSSTIWKIKSSKNSAANRELQKKLVKDNKKNRTKKK 120  
Db |||||-----VKDNKKNRTKKK 77  
QY 121 PTPKPPVVDAGSLDNGDFKVTTPDTSITQHNKVSTSPKITTAKPINRPSLPPNSDTS 180  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 78 PTPKPPVVDAGSLDNGDFKVTTPDTSITQHNKVSTSPKITTAKPINRPSLPPNSDTS 137  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 181 KETSLVNVKETTIVETKETTITNKQTSIDGKEKITTSAKETOSIEKTSAKDLAPTSKVLAKP 240  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 138 KETSLVNVKETTIVETKETTITNKQTSIDGKEKITTSAKETOSIEKTSAKDLAPTSKVLAKP 197  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 241 TPKAETTTKGPALTTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPT 300  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 198 TPKAETTTKGPALTTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPT 257  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 301 TPKEPAPTTTKEPAPTTTPKEPAPTTTKSAPTTPKEPAPTTTPKEPAPTTTPKEP 360  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 258 TPKEPAPTTTKEPAPTTTPKEPAPTTTKSAPTTPKEPAPTTTPKEPAPTTTPKEP 317  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 361 APPTTKEPTPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTT 420  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 318 APPTTKEPTPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTT 377  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 421 KEPSPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTPKEPAPTTTPKEP 480  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 378 KEPSPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTPKEPAPTTTPKEP 437  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 481 TPKKPAPTTTPKEPAPTTTPKEPAPTTTKKPAPAPKEPAPTTTPKETAPTTPKKLTPTTPEK 540  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 438 TPKKPAPTTTPKEPAPTTTPKEPAPTTTKKPAPAPKEPAPTTTPKETAPTTPKKLTPTTPEK 497  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 541 LAPTTPEKAPTTPEELAPTTPEEPTPTTPEEAPTTPKAAAPNTPKAPAPTTTPKEPAP 600  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 498 LAPTTPEKAPTTPEELAPTTPEEPTPTTPEEAPTTPKAAAPNTPKAPAPTTTPKEPAP 557  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 601 TPKEPAPTTTPKETAPTTPKGTATTTLKEPAPTTPKGPAPKEPAPTTTKEPTSTTSOKPAP 660  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 558 TPKEPAPTTTPKETAPTTPKGTATTTLKEPAPTTPKGPAPKEPAPTTTKEPTSTTSOKPAP 617  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 661 TTPKGATPTTPKEPAPTTTPKEPAPTTPKGTATTTLKEPAPTTPKGPAPKEPAPTTTKEPT 720  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 618 TTPKGATPTTPKEPAPTTTPKEPAPTTPKGTATTTLKEPAPTTPKGPAPKEPAPTTTKEPT 677  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 721 STTSOKPAPTTTPKETAPTTTPKEPAPTTTPKEPAPTTTKEPTSTTSOKPAPTTTKEPT 780  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 678 STTSOKPAPTTTPKETAPTTTPKEPAPTTTPKEPAPTTTKEPTSTTSOKPAPTTTKEPT 737  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 781 SPDESPELSAETPKALENSPKERGVPPTTKTAATKPEMTTITAKOKTTERDLRTTPEIT 840  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 738 SPDESPELSAETPKALENSPKERGVPPTTKTAATKPEMTTITAKOKTTERDLRTTPEIT 797  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 841 TAAPKMTKETATTTTEKTTESKITATTTQVTTSTTQDTPFKITTLKTTLLAPKVTTTKKT 900  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 798 TAAPKMTKETATTTTEKTTESKITATTTQVTTSTTQDTPFKITTLKTTLLAPKVTTTKKT 857  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 901 ITTTEIMNKBETAKPKDRATNSKATTPKQKPTKAPKKPTSTKKPTMPBRVRKPTTPT 960  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 858 ITTTEIMNKBETAKPKDRATNSKATTPKQKPTKAPKKPTSTKKPTMPBRVRKPTTPT 917  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 961 PRKMTSTMPBLNPTSRIAEAMLQTTTTRPNQTPNSKLVENVNPKSBDAGGAGETPHMLLRP 1020  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 918 PRKMTSTMPBLNPTSRIAEAMLQTTTTRPNQTPNSKLVENVNPKSBDAGGAGETPHMLLRP 977  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 1021 HVFMPEVTPDMVDYLPRVFNQGIINPMLS 1049  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 978 HVFMPEVTPDMVDYLPRVFNQGIINPMLS 1006  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Search completed: October 13, 2004, 11:59:01  
Job time : 24.2347 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 106.723 Seconds  
(without alignments)  
3171.696 Million cell updates/sec

Title: SEQ1-D

Perfect score: 5576

Sequence: 1 MAWKTLPIYILLLSVFVIQ.....DMDYLPRVFNQGIINPMLS 1049

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.1

Searched: 1356558 seqs, 322682953 residues

Total number of hits satisfying chosen parameters: 1356558

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA.\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US05\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	5576	100.0	1049	13	US-10-124-557-58	Sequence 58, Appl
2	5576	100.0	1313	13	US-10-124-557-142	Sequence 142, App
3	5561.9	99.7	1354	13	US-10-124-557-48	Sequence 48, Appl
4	5561.9	99.7	1363	13	US-10-124-557-52	Sequence 52, Appl
5	5546.9	99.5	1140	13	US-10-124-557-104	Sequence 104, App
6	5546.9	99.5	1404	9	US-09-802-207-30	Sequence 30, Appl
7	5546.9	99.5	1404	11	US-09-897-188-1	Sequence 1, Appli
8	5546.9	99.5	1404	13	US-10-124-557-2	Sequence 2, Appli
9	5546.9	99.5	1404	13	US-10-124-557-62	Sequence 62, Appl
10	5447.9	97.7	1314	13	US-10-124-557-50	Sequence 50, Appl
11	5393.3	96.7	1320	13	US-10-124-557-46	Sequence 46, Appl
12	5393.3	96.7	1320	13	US-10-124-557-60	Sequence 60, Appl
13	5379.2	96.5	1361	13	US-10-124-557-40	Sequence 40, Appl
14	5358.7	96.1	1038	13	US-10-124-557-74	Sequence 74, Appl

15 5358.7 96.1 1270 13 US-10-124-557-44 Sequence 44, Appl  
16 5344.6 95.9 1311 13 US-10-124-557-42 Sequence 42, Appl  
17 5235.9 93.9 1022 13 US-10-124-557-84 Sequence 84, Appl  
18 5011 89.9 941 13 US-10-124-557-14 Sequence 14, Appl  
19 2182.7 39.1 792 9 US-09-802-207-27 Sequence 27, Appl  
20 1409.9 25.3 538 14 US-10-038-694-3 Sequence 3, Appl  
21 1346.8 24.2 5179 9 US-09-922-217-1068 Sequence 1068, Appl  
22 1346.8 24.2 5179 9 US-09-833-263-1068 Sequence 1068, Appl  
23 1346.8 24.2 5179 13 US-10-025-380-1068 Sequence 1068, Appl  
24 1346.8 24.2 5179 16 US-10-734-564-121 Sequence 121, Appl  
25 1174.4 21.1 292 16 US-10-468-910-4 Sequence 4, Appl  
26 951.1 17.1 1460 14 US-10-295-037-428 Sequence 428, Appl  
27 931.8 16.7 1325 9 US-09-864-761-35612 Sequence 35612, A  
28 918 16.5 1367 9 US-09-801-368-108 Sequence 108, Appl  
29 860.6 15.4 3507 14 US-10-369-493-5784 Sequence 5784, Appl  
30 857.5 15.4 6642 14 US-10-369-493-5013 Sequence 5013, Appl  
31 855.3 15.3 2090 16 US-10-408-765A-2318 Sequence 2318, Appl  
32 834.9 15.0 3256 16 US-10-408-765A-174 Sequence 174, Appl  
33 834.9 15.0 3256 16 US-10-701-490-9 Sequence 9, Appl  
34 832.7 14.9 697 15 US-10-425-114-41545 Sequence 41545, A  
35 828.7 14.9 19723 15 US-10-084-846A-5 Sequence 5, Appl  
36 827.1 14.8 3256 9 US-09-919-172-98 Sequence 98, Appl  
37 827.1 14.8 3256 10 US-09-919-039-21 Sequence 21, Appl  
38 822.8 14.8 1151 10 US-09-825-751A-79 Sequence 79, Appl  
39 822 14.7 2397 15 US-10-282-122A-71232 Sequence 71232, A  
40 817.1 14.7 22152 16 US-10-715-066-5 Sequence 5, Appl  
41 816.4 14.6 5935 14 US-10-243-243A-8 Sequence 8, Appl  
42 816 14.6 5877 14 US-10-142-515-11 Sequence 11, Appl  
43 809.6 14.5 2971 14 US-10-146-473-50 Sequence 50, Appl  
44 799.7 14.3 1255 10 US-09-996-069-10 Sequence 10, Appl  
45 799.7 14.3 1255 14 US-10-171-311-158 Sequence 158, Appl

#### ALIGNMENTS

#### RESULT 1

US-10-124-557-58

; Sequence 58, Application US/10124557

; Publication No. US20020137894A1

; GENERAL INFORMATION:

APPLICANT: Turner, Katherine

Clark, Stephen C.

Jacobs, Kenneth

Hewick, Rodney M.

Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 CambridgePark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION NUMBER: US/10/124,557

FILING DATE: 16-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990

APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989

APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989



Query Match 100.0%; Score 5576; DB 13; Length 1313;  
Best Local Similarity 100.0%; Pred. No. 1.4e-139;  
Matches 1049; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAWKTLPIYLLLSVFIQVSSQELSCKGRFCFESPERGREGCDCAQCKKYDKCCPDYE 60  
DB 1 MAWKTLPIYLLLSVFIQVSSQELSCKGRFCFESPERGREGCDCAQCKKYDKCCPDYE 60  
QY 61 SFCAEHSVSENQESSSSSSSSSSSSSIWKIKSKNSAANRELQKLVKDNKNRTKKK 120  
DB 61 SFCAEHSVSENQESSSSSSSSSSSSSIWKIKSKNSAANRELQKLVKDNKNRTKKK 120  
QY 121 PTPKPPVVDAGSLDNGDFKVTTPDSTTTQHNKVSTSPKITTAKPINRPSLPNSDTS 180  
DB 121 PTPKPPVVDAGSLDNGDFKVTTPDSTTTQHNKVSTSPKITTAKPINRPSLPNSDTS 180  
QY 181 KETSLTVNKETTVETKETTNNKQSTDGKEKTSKETSIAKOLAPTSKVLAKP 240  
DB 181 KETSLTVNKETTVETKETTNNKQSTDGKEKTSKETSIAKOLAPTSKVLAKP 240  
QY 241 TPXAEITTKGPALTTKPEPTTTKBPASTTPKEPTTTIKSAPTTKBPATTTKSAPT 300  
DB 241 TPXAEITTKGPALTTKPEPTTTKBPASTTPKEPTTTIKSAPTTKBPATTTKSAPT 300  
QY 301 TPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 360  
DB 301 TPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 360  
QY 361 APPTKEPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 420  
DB 361 APPTKEPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 420  
QY 421 KEPSPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 480  
DB 421 KEPSPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 480  
QY 481 TPKKPAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 540  
DB 481 TPKKPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 540  
QY 541 LAPTTPEKAPTTPEELATTTPEPTTTPEPAPTTTKEPAPTTTKEPAPTTTKEP 600  
DB 541 LAPTTPEKAPTTPEELATTTPEPTTTPEPAPTTTKEPAPTTTKEPAPTTTKEP 600  
QY 601 TPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 660  
DB 601 TPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 660  
QY 661 TTPKGTAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 720  
DB 661 TTPKGTAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 720  
QY 721 STTSKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 780  
DB 721 STTSKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 780  
QY 781 SPDESTPELSABPTPKALENSKPEGVPTTKPAATKPEMTTAKDTERDLRTPTT 840  
DB 781 SPDESTPELSABPTPKALENSKPEGVPTTKPAATKPEMTTAKDTERDLRTPTT 840  
QY 841 TPAOKMTKETAATTTKTTESKITATTQVTSSTTTQDTPFKITLTKTTTAPKVTTKKT 900  
DB 841 TPAOKMTKETAATTTKTTESKITATTQVTSSTTTQDTPFKITLTKTTTAPKVTTKKT 900  
QY 901 ITTTEIMNKPESTAKPDRATNSKATTPKQKPTKAPKPTSTKPKPMRVRKPKTTPT 960  
DB 901 ITTTEIMNKPESTAKPDRATNSKATTPKQKPTKAPKPTSTKPKPMRVRKPKTTPT 960  
QY 961 PRKMTSTWPELNPSTRIAEAMIQTTTRPNQTPNSKLVEVNPKSEDAGGEGETHMLLRP 1020  
DB 961 PRKMTSTWPELNPSTRIAEAMIQTTTRPNQTPNSKLVEVNPKSEDAGGEGETHMLLRP 1020  
QY 1021 HVFMPEVTPDMDYLPVFNQGIINPMLS 1049

DB 1021 HVFMPEVTPDMDYLPVFNQGIINPMLS 1049

RESULT 3  
US-10-124-557-48  
; Sequence 48, Application US/10124557  
; Publication No. US20020137894A1  
; GENERAL INFORMATION:  
; APPLICANT: Turner, Katherine  
; Clark, Stephen C.  
; Jacobs, Kenneth  
; Hewick, Rodney M.  
; Gesner, Thomas G.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 Cambridgepark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10124,557  
; FILING DATE: 16-Apr-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 876-1170  
; TELEFAX: (617) 876-5851  
; INFORMATION FOR SEQ ID NO: 48:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1354 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 48:  
US-10-124-557-48

Query Match 99.7%; Score 5561.9; DB 13; Length 1354;  
Best Local Similarity 96.2%; Pred. No. 3.4e-139;  
Matches 1049; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAWKTLPIYLLLSVFIQVSSQ----- 25  
DB 1 MAWKTLPIYLLLSVFIQVSSQDLSSCAGRCGEGYSDRATCNCYNCQHYMECCPDF 60  
QY 26 -----ELSCKGRFCFESPERGREGCDCAQCKKYDKCCPDYEFCAEHSVSENQESSSS 79  
DB 61 KRVTAELSCKGRFCFESPERGREGCDCAQCKKYDKCCPDYEFCAEHSVSENQESSSS 120  
QY 80 SSSSSSSSIWKIKSKNSAANRELQKLVKDNKNRTKKKPTKPPVVDAGSLDNGD 139  
DB 121 SSSSSSSSIWKIKSKNSAANRELQKLVKDNKNRTKKKPTKPPVVDAGSLDNGD 180

Qy	140	FKVTTDSTTQHNKVSTSPKII	TTAKPIINPRPSLPNSDTSKETS	UTNVNKETIVETKET	199
Db	181	FKVTTDSTTQHNKVSTSPKII	TTAKPIINPRPSLPNSDTSKETS	UTNVNKETIVETKET	240
Qy	200	TTNKQSTSDGKEKITTSAKETQS	TEKTSAKDLAPTSKVLAKPTKAB	TTTTKGGALTTTPKEP	259
Db	241	TTNKQSTSDGKEKITTSAKETQS	TEKTSAKDLAPTSKVLAKPTKAB	TTTTKGGALTTTPKEP	300
Qy	260	TPPTPKEPASTTTPKEBPTPI	TKSAPTTTPKEPATTTKSAPTT	TPKBPATTTTKEPAPTTTPK	319
Db	301	TPPTPKEPASTTTPKEBPTPI	TKSAPTTTPKEPATTTKSAPTT	TPKBPATTTTKEPAPTTTPK	360
Qy	320	EPAPTTTKEPAPTTTTSAPTT	PKKPAPTTTPKEPATTTPKBPAT	TTTTKEBPTPTTPKEPAPT	379
Db	361	EPAPTTTKEPAPTTTTSAPTT	PKKPAPTTTPKEPATTTPKBPAT	TTTTKEBPTPTTPKEPAPT	420
Qy	380	TKBPAPTTTKEPAPTAAPKBP	ATTTTPKEPATTTPKEPATTTKBP	TTTTTKEPSTTPKEPAPTTTKGA	439
Db	421	TKBPAPTTTKEPAPTAAPKBP	ATTTTPKEPATTTPKEPATTTKBP	TTTTTKEPSTTPKEPAPTTTKGA	480
Qy	440	PTTTTKEPAPTTTTSAPTT	PKBPSTTTTKEPATTTPKEPATT	TPKBPATTTTKEPAPTTTPK	499
Db	481	PTTTTKEPAPTTTTSAPTT	PKBPSTTTTKEPATTTPKEPATT	TPKBPATTTTKEPAPTTTPK	540
Qy	500	EPAPTTTTPKAPAPKBPATTT	PKETAPTTTPKKLTPTTPPKLAP	TTTPPKAPTTTPKEPAPTTPEELAP	559
Db	541	EPAPTTTTPKAPAPKBPATTT	PKETAPTTTPKKLTPTTPPKLAP	TTTPPKAPTTTPKEPAPTTPEELAP	600
Qy	560	TTTTPEEPPTTTPEBPAPTT	PKAAAPNTPKEPAPTTTPKEPATT	TPKBPATTTPKETAPTTTPK	619
Db	601	TTTTPEEPPTTTPEBPAPTT	PKAAAPNTPKEPAPTTTPKEPATT	TPKBPATTTPKETAPTTTPK	660
Qy	620	GTAPTTTKEPAPTTPKPAPKEL	APTTTTPKBPSTTSDKPAPTTPKG	ATPTTPKEPAPTTTP	679
Db	661	GTAPTTTKEPAPTTPKPAPKEL	APTTTTPKBPSTTSDKPAPTTPKG	ATPTTPKEPAPTTTP	720
Qy	680	KBAPATTPKGTAPTTTKEPAPT	TPKBPAPKELAPTTTTPKGSTTS	DKPAPTTTPKEAPITT	739
Db	721	KBAPATTPKGTAPTTTKEPAPT	TPKBPAPKELAPTTTTPKGSTTS	DKPAPTTTPKEAPITT	780
Qy	740	PKBPATTPKBPATTPETPPT	SEVSTTPTTTKEPTTIHKSPDES	TEPBUSAPETPKALE	799
Db	781	PKBPATTPKBPATTPETPPT	SEVSTTPTTTKEPTTIHKSPDES	TEPBUSAPETPKALE	840
Qy	800	NSPKPBGVPTTKTPAAATKPE	MTTITAKDKTTTERDLRTTPTTAA	PAKMTKETATTTTEKTT	859
Db	841	NSPKPBGVPTTKTPAAATKPE	MTTITAKDKTTTERDLRTTPTTAA	PAKMTKETATTTTEKTT	900
Qy	860	SKITATTTQVSTTTQDTPPK	ITITLKITTLAPKVTTTXXIIITTE	IMNKPEETAKPKDR	919
Db	901	SKITATTTQVSTTTQDTPPK	ITITLKITTLAPKVTTTXXIIITTE	IMNKPEETAKPKDR	960
Qy	920	ATNSKATTPKPKETKAPKPK	PTSCKPKTPVRNPKKPTTTPRKMT	STMPELNPTSRIAE	979
Db	961	ATNSKATTPKPKETKAPKPK	PTSCKPKTPVRNPKKPTTTPRKMT	STMPELNPTSRIAE	1020
Qy	980	AMLQTTTRPNQTPNSKLVE	NPKSGEDAGGAGETPHMLLR	PHVEMPEVTTDMDYLPRVPN	1039
Db	1021	AMLQTTTRPNQTPNSKLVE	NPKSGEDAGGAGETPHMLLR	PHVEMPEVTTDMDYLPRVPN	1080
Qy	1040	QGIINPMLS 1049			
Db	1081	QGIINPMLS 1090			

RESULT 4	
US-10-124-557-52	
; Sequence 52, Application US/10124557	
; Publication No. US20020137894A1	
; GENERAL INFORMATION:	
; APPLICANT: Turner, Katherine	
; Clark, Stephen C.	
; Jacobs, Kenneth	





```

RESULT 7
US-09-897-188-1
; Sequence 1, Application US/09897188
; Publication No. US20040072741A1
; GENERAL INFORMATION:
; APPLICANT: Jay, Gregory D.
; TITLE OF INVENTION: Tribonectin Polypeptides and Uses Thereof
; FILE REFERENCE: 21486-026 CIP2
; CURRENT APPLICATION NUMBER: US/09/897,188
; CURRENT FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: 09/298,970
; PRIOR FILING DATE: 1999-04-23
; PRIOR APPLICATION NUMBER: 09/556,246
; PRIOR FILING DATE: 2000-04-24
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1404
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-897-188-1

Query Match          99.5%; Score 5546.9; DB 11; Length 1404;
Best Local Similarity 92.0%; Pred. No. 8.7e-139;
Matches 1049; Conservative 0; Mismatches 0; Indels 91; Gaps 2;

QY 1 MAWKTLPIYLLLLLSVFVIQVSSQ----- 25
DB 1 MAWKTLPIYLLLLLSVFVIQVSSQ----- 60
QY 26 -----ELSKGRCFERSFERGECDDAOCKKYDKCPOYESFCA----- 64
DB 61 KRVCTAELSKGRCFERSFERGECDDAOCKKYDKCPOYESFCAEVHNPTSPSSKKAP 120
QY 65 -----BEHSVSENQESSSSSSSSSSSTI 89
DB 121 PPSGASQTIKSTTKRSPKPNKKTKKVIIESEITEHSVSENQESSSSSSSSSTI 180
QY 90 KIKSSKNSAANRELOKLVKONKKNRTKKTTPKPPVVDVDEAGSLDNGDFKVTTPDTST 149
DB 181 KIKSSKNSAANRELOKLVKONKKNRTKKTTPKPPVVDVDEAGSLDNGDFKVTTPDTST 240
QY 150 TQHNKVSTSPKITTAKPINRPSLPNSDTSKETSUTVNKETTIVETKETTNNKQSTDG 209
DB 241 TQHNKVSTSPKITTAKPINRPSLPNSDTSKETSUTVNKETTIVETKETTNNKQSTDG 300
QY 210 KBKTTSAKETQSIEKTSADLAPTSKVLAKPTPKAFTTTKGPALITPKKEPTTTPKEPAS 269
DB 301 KEKTTSAKETQSIEKTSADLAPTSKVLAKPTPKAFTTTKGPALITPKKEPTTTPKEPAS 360
QY 270 TTPKEPTTPTTKSAPTTKKEPAPTTTKSAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEP 329
DB 361 TTPKEPTTPTTKSAPTTKKEPAPTTTKSAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEP 420
QY 330 APTTTKSAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTK 389
DB 421 APTTTKSAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTK 480
QY 390 EPAPTAPKPAATTPKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTT 449
DB 481 EPAPTAPKPAATTPKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTT 540
QY 450 TTKSAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEP 509
DB 541 TTKSAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEP 600
QY 510 APTAPKEPAPTTPKETAPTTPKLTPPTTPEKLAPTTPEKLAPTTPEKLAPTTPEEPTPTT 569
DB 601 APTAPKEPAPTTPKETAPTTPKLTPPTTPEKLAPTTPEKLAPTTPEEPTPTT 660
QY 570 PEEAPPTPKAAAPNTPKEPAPTTPKETAPTTPKETAPTTPKETAPTTPKETAPTTPKETAPTT 629

```

## RESULT 8

```

US-10-124-557-2
; Sequence 2, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luan
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:

```







Best Local Similarity 92.0%; Pred. No. 8.7e-139; Matches 1049; Conservative 0; Mismatches 0; Indels 91; Gaps 2;									
QY	1	MAWKTLPIYLLLLSVFVIQOVSSQ-----	25						
Db	1	MAWKTLPIYLLLLSVFVIQOVSSQDLSSCAGRCGEGYGRDATCNCYDNCQHYMECCPDF	60						
QY	26	-----ELSCGRCFESFERGECDCDAOCKKYDKCCPDYEFCA-----	64						
Db	61	KEVCTAELSCGRCFESFERGECDCDAOCKKYDKCCPDYEFCAEVENPTSPPSKKAP	120						
QY	65	-----EHSVSVENQESSSSSSSSSSSTIIM	89						
Db	121	PPSGASQTIKSTTKRSPKPNKKTKKVTESEITEHSVSVENQESSSSSSSSSSSTIIM	180						
QY	90	KIKSSKNSAANELOKKLVKDNKKNRKTKKTPPPVVDVDEAGSLDNGDFKVTTPDST	149						
Db	181	KIKSSKNSAANELOKKLVKDNKKNRKTKKTPPPVVDVDEAGSLDNGDFKVTTPDST	240						
QY	150	TOHNKVSTSPKITTAKPINPRESLPNSDTSKETSITVNKETTIVETKETTINKOTSTDG	209						
Db	241	TOHNKVSTSPKITTAKPINPRESLPNSDTSKETSITVNKETTIVETKETTINKOTSTDG	300						
QY	210	KEKTSKAKETQSIKTSKOLAPTSKVLAKPTPKAETTTKGPALTTPKETPTTTPKEPAS	269						
Db	301	KEKTSKAKETQSIKTSKOLAPTSKVLAKPTPKAETTTKGPALTTPKETPTTTPKEPAS	360						
QY	270	TTPKEPTTTIKSAPTTKEPATTTTKSAPTTKEPATTTKEPATTTKEPATTTKEP	329						
Db	361	TTPKEPTTTIKSAPTTKEPATTTTKSAPTTKEPATTTKEPATTTKEPATTTKEP	420						
QY	330	APTPTKSAPTTPKEPATTPPKKPAPTTPKPATTPKEPATTPKEPATTTKEPATTPK	389						
Db	421	APTPTKSAPTTPKEPATTPPKKPAPTTPKPATTPKEPATTPKEPATTTKEPATTPK	480						
QY	390	EPAPTAPKAPAPTTKEPATTPKEPATTTKEPATTTKEPATTTKSAPTTTKEPAT	449						
Db	481	EPAPTAPKAPAPTTKEPATTPKEPATTTKEPATTTKEPATTTKSAPTTTKEPAT	540						
QY	450	TTKSAPTTKESPTTTKEPATTTKEPATTTKEPATTTKEPATTTKEPATTTKEP	509						
Db	541	TTKSAPTTKESPTTTKEPATTTKEPATTTKEPATTTKEPATTTKEPATTTKEP	600						
QY	510	APTAPKEPATTPKETAPTTPKLTPTPEKLAPTTPEKLAPTTPEELAPTTPEEPTTT	569						
Db	601	APTAPKEPATTPKETAPTTPKLTPTPEKLAPTTPEKLAPTTPEELAPTTPEEPTTT	660						
QY	570	PEEPAPTTKAAAPNTKEPATTTKEPATTTKEPATTTKEPATTTKEPATTTKEP	629						
Db	661	PEEPAPTTKAAAPNTKEPATTTKEPATTTKEPATTTKEPATTTKEPATTTKEP	720						
QY	630	APTTPKEAPKELAPTTKEPTSTSDKAPATTPKGATTPKEPATTTKEPATTTKEP	689						
Db	721	APTTPKEAPKELAPTTKEPTSTSDKAPATTPKGATTPKEPATTTKEPATTTKEP	780						
QY	690	TAPTLKEPATTPKAPKELAPTTKGTSTSDKAPATTPKETAPTTPKEPATTTKEP	749						
Db	781	TAPTLKEPATTPKAPKELAPTTKGTSTSDKAPATTPKETAPTTPKEPATTTKEP	840						
QY	750	KPATTPPTTSEVSTPTTTKEPTTIHKSPPDESTPELSAETTPKALENSKPEGVP	809						
Db	841	KPATTPPTTSEVSTPTTTKEPTTIHKSPPDESTPELSAETTPKALENSKPEGVP	900						
QY	810	TKTPAATPEMTTAKKTTEDRLTETTTAAKMTKETATTTTEKTESKITATTQV	869						
Db	901	TKTPAATPEMTTAKKTTEDRLTETTTAAKMTKETATTTTEKTESKITATTQV	960						
QY	870	TSTTTQDTPPKITLTKITTLAPKVTTTKITTTTTEIMNKPEETAAPKORATNSKATPK	929						
Db	961	TSTTTQDTPPKITLTKITTLAPKVTTTKITTTTTEIMNKPEETAAPKORATNSKATPK	1020						
QY	930	POKPTKAPKPTSTTKPKTMPRVKPTTTPRKMSTMPELNPTSRIAEAMLOTTTRPN	989						

1021 POKPTKAPKPTSTTKPKTMPRVKPTTTPRKMSTMPELNPTSRIAEAMLOTTTRPN 1080

990 QTPNSKLVFNPKSBDAGAGETPHMLLRPHVFWPEVTPDMYDLPRVFNQGIINPMLS 1049

1081 QTPNSKLVFNPKSBDAGAGETPHMLLRPHVFWPEVTPDMYDLPRVFNQGIINPMLS 1140

RESULT 10

US-10-124-557-50

Sequence 50, Application US/10124557

Publication NO. US20020137894A1

GENERAL INFORMATION:

APPLICANT: Turner, Katherine

Clark, Stephen C.

Jacobs, Kenneth

Hewick, Rodney M.

Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 CambridgePark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/124,557

FILING DATE: 16-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990

APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989

APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Cseirr, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)876-1170

TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 50:

SEQUENCE CHARACTERISTICS:

LENGTH: 1314 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 50:

US-10-124-557-50

Query Match

Best Local Similarity 97.7%; Score 5447.9; DB 13; Length 1314;

Matches 1027; Conservative 7; Mismatches 15; Indels 1; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQOVSSQEL-SCKGRCFESFERGECDCDAOCKKYDKCCPDF 59

Db 1 MAWKTLPIYLLLLSVFVIQOVSSQDLSSCAGRCGEGYGRDATCNCYDNCQHYMECCPDF 60

QY 60 ESFCAEHSVSVENQESSSSSSSSSSSTIIMKIKSSKNSAANELOKKLVKDNKKNRKTK 119

Db 61 KRVTCTAHSVSVENQESSSSSSSSSSSTIIMKIKSSKNSAANELOKKLVKDNKKNRKTK 120

QY 120 KPTPKPPVVDVDEAGSLDNGDFKVTTPDTSITQHNKVSTSPKITTAKPINPREPSLPNSDT 179

Db 121 KPTKPPVWDAGSLDNGDFKVTTPDSTTOHNVSTSPKITTAKPINRPSLPPNSDT 180  
QY 180 SKETSLTVNKEVTVETKTTTTNKTSTGKEKTTSAKETQSIKTSKOLAPTSKVLAK 239  
Db 181 SKETSLTVNKEVTVETKTTTTNKTSTGKEKTTSAKETQSIKTSKOLAPTSKVLAK 240  
QY 240 PTPKAEITTKGPAITTKPEPTTTKPEASTTKPEPTTTIKSAPTTPKEPATTTKSAP 299  
Db 241 PTPKAEITTKGPAITTKPEPTTTKPEASTTKPEPTTTIKSAPTTPKEPATTTKSAP 300  
QY 300 TTPKEPATTTKEPATTTKEPATTTKSAPTTTKSAPTTTKEPATTTKPKPATTPKE 359  
Db 301 TTPKEPATTTKEPATTTKEPATTTKSAPTTTKSAPTTTKEPATTTKPKPATTPKE 360  
QY 360 PAPTTPKEPTTPTPKEPATTTKEPATTTKPAATPKKPAATTPKEPATTTKEPATTT 419  
Db 361 PAPTTPKEPTTPTPKEPATTTKEPATTTKPAATPKKPAATTPKEPATTTKEPATTT 420  
QY 420 TKEPSPTTPKEPATTTKSAPTTTKEPATTTKSAPTTTKSAPTTTKSAPTTTKEPAT 479  
Db 421 TKEPSPTTPKEPATTTKSAPTTTKEPATTTKSAPTTTKSAPTTTKSAPTTTKEPAT 480  
QY 480 TTPKKEPATTTKEPATTTKKEPATTTKKEPATTTKKEPATTTKKEPATTTKKEPAT 539  
Db 481 TTPKKEPATTTKEPATTTKKEPATTTKKEPATTTKKEPATTTKKEPATTTKKEPAT 540  
QY 540 KLAAPTTPKEPATTTPEELAPTTPPEPTTPPEPATTTPKAAANTPKPEPATTTKEPAT 599  
Db 541 KLAAPTTPKEPATTTPEELAPTTPPEPTTPPEPATTTPKAAANTPKPEPATTTKEPAT 600  
QY 600 TTPKEPATTTKETAPTTPKGTAPTTLKKEPATTTKKEPATTTKKEPATTTKKEPAT 659  
Db 601 TTPKEPATTTKETAPTTPKGTAPTTLKKEPATTTKKEPATTTKKEPATTTKKEPAT 660  
QY 660 PTPPKGTAPTTPKEPATTTKGTAPTTLKKEPATTTKKEPATTTKKEPATTTKKEPAT 719  
Db 661 PTPPKGTAPTTPKEPATTTKGTAPTTLKKEPATTTKKEPATTTKKEPATTTKKEPAT 720  
QY 720 TSTTSKDPAPTTPKETAPTTPKEPATTTKKEPATTTKKEPATTTKKEPATTTKKEPAT 779  
Db 721 TSTTSKDPAPTTPKETAPTTPKEPATTTKKEPATTTKKEPATTTKKEPATTTKKEPAT 780  
QY 780 KSPDSTPELSAEPKALENSPKEGVPTTKTPAATKPEMTTTAKDKTTERDLRTTPT 839  
Db 781 KSPDSTPELSAEPKALENSPKEGVPTTKTPAATKPEMTTTAKDKTTERDLRTTPT 840  
QY 840 TTAAPKMTKETATTTETKTTESKITATTTQVSTSTTQDITTPFKITLTKTTTLAPKVTITKK 899  
Db 841 TTAAPKMTKETATTTETKTTESKITATTTQVSTSTTQDITTPFKITLTKTTTLAPKVTITKK 900  
QY 900 TTTTTEIMNKPEATAKPKORATNSKATTPKQPKPKAPKPKPTSTKPKMTMPVRPKKTP 959  
Db 901 TTTTTEIMNKPEATAKPKORATNSKATTPKQPKPKAPKPKPTSTKPKMTMPVRPKKTP 960  
QY 960 TPKMTSTPELNPSTSRIAEMLQTTTRNQTPNSKLVFNPKSDEAGGAEGETPHMLR 1019  
Db 961 TPKMTSTPELNPSTSRIAEMLQTTTRNQTPNSKLVFNPKSDEAGGAEGETPHMLR 1020  
QY 1020 PHVFMEPVTPDMDYLRPVNQGIINPMLS 1049  
Db 1021 PHVFMEPVTPDMDYLRPVNQGIINPMLS 1050

RESULT 11  
US-10-124-557-46  
; Sequence 46, Application US/10124557  
; Publication No. US20020137894A1  
; GENERAL INFORMATION:  
; APPLICANT: Turner, Katherine  
; Clark, Stephen C.  
; Jacobs, Kenneth  
; Hewick, Rodney M.

; Gesner, Thomas G.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 Cambridgepark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/124,557  
; FILING DATE: 16-Apr-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Juann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 46:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1320 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 46:  
US-10-124-557-46  
  
Query Match 96.7%; Score 5393.3; DB 13; Length 1320;  
Best Local Similarity 96.1%; Pred. No. 9.3e-135;  
Matches 1015; Conservative 8; Mismatches 26; Indels 7; Gaps 1;  
  
QY 1 MAWKTLPIYLLLLSVFVIQQVSSQELSCGRCFCFSFERGRCDCDAQCKYDKCCPDYE 60  
Db 1 MAWKTLPIYLLLLSVFVIQQVSSQELSCGRCFCFSFERGRCDCDAQCKYDKCCPDYE 60  
QY 61 SFCAEBHVSVENQESSSSSSSSSSSTIWKIKSSKNSAANRELQKL-----KVKDNK 113  
Db 61 SFCAEBHVSVENQESSSSSSSSSSSTIWKIKSSKNSAANRELQKL-----KVKDNK 120  
QY 114 KNRTRKKKPTPKPPVVDVDEAGSLDNGDFKVTTPDSTTOHNVSTSPKITTAKPINRPSL 173  
Db 121 KNRTRKKKPTPKPPVVDVDEAGSLDNGDFKVTTPDSTTOHNVSTSPKITTAKPINRPSL 180  
QY 174 PPNSDTSKETSLSLVNKEVTVETKTTTTNKTSTGKEKTTSAKETQSIKTSKOLAPT 233  
Db 181 PPNSDTSKETSLSLVNKEVTVETKTTTTNKTSTGKEKTTSAKETQSIKTSKOLAPT 240  
QY 234 SKVLAKPTPKAEITTKGPAITTKPEPTTTKPEASTTKPEPTTTIKSAPTTPKEPAT 293  
Db 241 SKVLAKPTPKAEITTKGPAITTKPEPTTTKPEASTTKPEPTTTIKSAPTTPKEPAT 300  
QY 294 TTKSAPTTPKEPATTTKKEPATTTKKEPATTTKKEPATTTKSAPTTTKKEPATTTKPKPA 353  
Db 301 TTKSAPTTPKEPATTTKKEPATTTKKEPATTTKKEPATTTKSAPTTTKKEPATTTKPKPA 360  
QY 354 PTTTPKEPATTTKPEPTTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTPK 413



Db 601 PKPAPPTPKGAPPTPKETAPTTKGTAPTTLKKEPAPTTKPKAPKELAPTTTKEPTST 660  
QY 654 TSDKAPPTPKGTAPTTKKEPAPTTKGPAPTTKGTAPTTLKKEPAPTTKPKAPKELAP 713  
Db 661 TSDKAPPTPKGTAPTTKKEPAPTTKGPAPTTKGTAPTTLKKEPAPTTKPKAPKELAP 720  
QY 714 TTTKGTSTSDKAPPTPKETAPTTKKEPAPTTKGPAPTTKGPAPTTKGPAPTTTSEVSTTTK 773  
Db 721 TTTKGTSTSDKAPPTPKETAPTTKKEPAPTTKGPAPTTKGPAPTTTSEVSTTTK 780  
QY 774 EPTTHKSDSTPELSAETPKALENSPKGVPPTTKPAATKPEMTTAKDKTERDL 833  
Db 781 EPTTHKSDSTPELSAETPKALENSPKGVPPTTKPAATKPEMTTAKDKTERDL 840  
QY 834 RTTPTTTAAAPKMTKETAATTTKTESKITATTTQVSTTTTQDTPPKITTLKTTTLPK 893  
Db 841 RTTPTTTAAAPKMTKETAATTTKTESKITATTTQVSTTTTQDTPPKITTLKTTTLPK 900  
QY 894 VTTTKKTTTITIMNKPEETAKPKORATNSKATTPKPKETKAPKPTSTKPKTMPVR 953  
Db 901 VTTTKKTTTITIMNKPEETAKPKORATNSKATTPKPKETKAPKPTSTKPKTMPVR 960  
QY 954 KPRTTTPRKMTSTMPELNPTSRIAEAMLTQTTTRPNQTPNSKLVEVNPKSEDAAGGAEGT 1013  
Db 961 KPRTTTPRKMTSTMPELNPTSRIAEAMLTQTTTRPNQTPNSKLVEVNPKSEDAAGGAEGT 1020  
QY 1014 PHEMLRPHVMEVTPDMDYLPRVPNQGIINPMLS 1049  
Db 1021 PHEMLRPHVMEVTPDMDYLPRVPNQGIINPMLS 1056

RESULT 13  
US-10-124-557-40  
; Sequence 40, Application US/10124557  
; Publication No. US20020137894A1  
; GENERAL INFORMATION:  
; APPLICANT: Turner, Katherine  
; Clark, Stephen C.  
; Jacobs, Kenneth  
; Hewick, Rodney M.  
; Gesner, Thomas G.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/124,557  
; FILING DATE: 16-Apr-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JUN-1991  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 40:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1361 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 40:  
US-10-124-557-40  
  
Query Match 96.5%; Score 5379.2; DB 13; Length 1361;  
Best Local Similarity 92.5%; Pred. No. 2.3e-134;  
Matches 1015; Conservative 8; Mismatches 26; Indels 48; Gaps 2;  
  
QY 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25  
Db 1 MAWKTLPIYLLLLSVFVIQVSSQDLSCAGRCGEGYSDATCNDYNCQHVNECCPDF 60  
QY 26 -----ELSCKGRCTESFERGECDCDAQCKKYDKCCPDYSEFCABEHSVSENQESSSS 79  
Db 61 KRVCTAELSCKGRCTESFERGECDCDAQCKKYDKCCPDYSEFCABEHSVSENQESSSS 120  
QY 80 SSSSSSTTIWKSSKNSAANRELQKL-----KVKNKKKRTKKKTPKPPVVDK 132  
Db 121 PPSGASQTIKSTTKRSPKPPNKKTKKVISEBEITEVKDKKKRTKKKTPKPPVVDK 180  
QY 133 SGLDNGDFKVTTPDSTTQHNVKSTSPKLTAKPINRPSLPNSDTSKETSLSLVNKEIT 192  
Db 181 SGLDNGDFKVTTPDSTTQHNVKSTSPKLTAKPINRPSLPNSDTSKETSLSLVNKEIT 240  
QY 193 VETKETTTNKOTS TDGKEKTTSAKETQSIETSAKDLPSTSKVLAKPTPKAETTTK 252  
Db 241 VETKETTTNKOTS TDGKEKTTSAKETQSIETSAKDLPSTSKVLAKPTPKAETTTK 300  
QY 253 LTTTKEPTTTPKEPASTTTPKEPTTIKSAPTTTPKEPAPTTTTPKEPAPTTTTPKE 312  
Db 301 LTTTKEPTTTPKEPASTTTPKEPTTIKSAPTTTPKEPAPTTTTPKEPAPTTTTPKE 360  
QY 313 PAPTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKE 372  
Db 361 PAPTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKE 420  
QY 373 PKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKE 432  
Db 421 PKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKE 480  
QY 433 PTTTKGAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKE 492  
Db 481 PTTTKGAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKE 540  
QY 493 PAPTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKE 552  
Db 541 PAPTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKE 600  
QY 553 TPBELAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKE 612  
Db 601 TPBELAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKE 660  
QY 613 TAPTTPKGTAAPTTLKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKE 672  
Db 661 TAPTTPKGTAAPTTLKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKE 720  
QY 673 EPAPTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKE 732  
Db 721 EPAPTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKE 780  
QY 733 KETAPTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKE 792  
Db 781 KETAPTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKE 840  
QY 793 PTPKALENSPKGVPPTTKPAATKPEMTTAKDKTERDLRTTPTTTTAAAPKMTKETAT 852

Db 841 PTPKALNSKPGVPTTKTAAKPKPMTTAKDKTERDLRTTPETTTAAKPKMTKETAT 900  
QY 853 TTEKTTESKITATTTOVTSSTTQDTPFKITLTKTTTLAPKVTTKTKTITTIMNKPEE 912  
Db 901 TTEKTTESKITATTTOVTSSTTQDTPFKITLTKTTTLAPKVTTKTKTITTIMNKPEE 960  
QY 913 TAKPKDRATNSKATTPKPKQPTKAPKPKPTSTKKPKTMPRVRKPKTTPTRKMTSTMPELN 972  
Db 961 TAKPKDRATNSKATTPKPKQPTKAPKPKPTSTKKPKTMPRVRKPKTTPTRKMTSTMPELN 1020  
QY 973 PTSRIAEAMLQTTTRPNQTPNSKLVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMD 1032  
Db 1021 PTSRIAEAMLQTTTRPNQTPNSKLVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMD 1080  
QY 1033 YLPRVENQGIINPMLS 1049  
Db 1081 YLPRVENQGIINPMLS 1097

RESULT 14  
US-10-124-557-74  
; Sequence 74, Application US/10124557  
; Publication No. US20020137894A1  
; GENERAL INFORMATION:  
; APPLICANT: Turner, Katherine  
; Clark, Stephen C.  
; Jacobs, Kenneth  
; Hewick, Rodney M.  
; Gesner, Thomas G.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSER: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/124,557  
; FILING DATE: 16-Apr-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: G1 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 74:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1038 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 74:

US-10-124-557-74

Query Match 96.1%; Score 5358.7; DB 13; Length 1038;  
Best Local Similarity 95.9%; Pred. No. 5.8e-134;  
Matches 1006; Conservative 0; Mismatches 0; Indels 43; Gaps 1;  
QY 1 MAWKTLPIYLLLLLSVFVIQQVSSQELSCKGRCFESFERGREGDCDCAQCKYDKCCPDYE 60  
Db 1 MAWKTLPIYLLLLLSVFVIQQVSSQELSCKGRCFESFERGREGDCDCAQCKYDKCCPDYE 60  
QY 61 SFCAEHSVSENQESSSSSSSSSSSIWKIKSSKNSAANRELQKKLVKXDNKNRRTKK 120  
Db 61 SFCAE-----VKDNKNRRTKK 77  
QY 121 PTPKPPVVDAGSLGNDPKVTPDTSSTTOHNVKSTSPKITTAKINPRPSLPNNSDTS 180  
Db 78 PTPKPPVVDAGSLGNDPKVTPDTSSTTOHNVKSTSPKITTAKINPRPSLPNNSDTS 137  
QY 181 KETSLTVNKETTIVETKETTNNKQSTDGKEKTTSAKETQSIEKTSAKDLAPTSKVLAKP 240  
Db 138 KETSLTVNKETTIVETKETTNNKQSTDGKEKTTSAKETQSIEKTSAKDLAPTSKVLAKP 197  
QY 241 TPKAETTTKGPALTTTPKEPTTPPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPT 300  
Db 198 TPKAETTTKGPALTTTPKEPTTPPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPT 257  
QY 301 TPKEPAPTTTKEPAPTTTPKEPAPTTTKSAPTTKSAPTTPKEPAPTTTPKAPATTPKEP 360  
Db 258 TPKEPAPTTTKEPAPTTTPKEPAPTTTKSAPTTKSAPTTPKEPAPTTTPKAPATTPKEP 317  
QY 361 APPTTPKEPTTPPKEPAPTTTPKEPAPTTAPKAPATTPKAPATTPKEPAPTTTPKEPAPTTT 420  
Db 318 APPTTPKEPTTPPKEPAPTTTPKEPAPTTAPKAPATTPKAPATTPKEPAPTTTPKEPAPTTT 377  
QY 421 KEPSPTTPKEPAPTTTKSAPTTKSAPTTKSAPTTPKEPSPTTKKEPAPTTTPKEPAP 480  
Db 378 KEPSPTTPKEPAPTTTKSAPTTKSAPTTKSAPTTPKEPSPTTKKEPAPTTTPKEPAP 437  
QY 481 TPKKAPATTPKEPAPTTTPKEPAPTTTKKAPATAPKAPATTPKETAATTPKLTTPPTPEK 540  
Db 438 TPKKAPATTPKEPAPTTTPKEPAPTTTKKAPATAPKAPATTPKETAATTPKLTTPPTPEK 497  
QY 541 LAPPTPEKAPATTPBELAPTTPEEPTTPPEPAPTTPKAAAPNTPKAPATTPKEPAP 600  
Db 498 LAPPTPEKAPATTPBELAPTTPEEPTTPPEPAPTTPKAAAPNTPKAPATTPKEPAP 557  
QY 601 TPKEPAPTTPKETAATTPKGTAPATTLKEPAPTTPKKAPKELAPTTTKEPTSTTSKAP 660  
Db 558 TPKEPAPTTPKETAATTPKGTAPATTLKEPAPTTPKKAPKELAPTTTKEPTSTTSKAP 617  
QY 661 TTPKGTAPTTTPKEPAPTTTPKEPAPTTPKGTATTTIKEPAPTTPKKAPKELAPTTTKGPT 720  
Db 618 TTPKGTAPTTTPKEPAPTTTPKEPAPTTPKGTATTTIKEPAPTTPKKAPKELAPTTTKGPT 677  
QY 721 STTSDKAPATTPPKETAATTPKEPAPTTPKKAPATTPPEPTTSEVSTPTTKKEPTTIHK 780  
Db 678 STTSDKAPATTPPKETAATTPKEPAPTTPKKAPATTPPEPTTSEVSTPTTKKEPTTIHK 737  
QY 781 SPDESTPELSAEPPTKALENSKPEGVPTTKPAATKPEMTTTAKDKTERDLRTTPETT 840  
Db 738 SPDESTPELSAEPPTKALENSKPEGVPTTKPAATKPEMTTTAKDKTERDLRTTPETT 797  
QY 841 TAAPKMTKETATTTTEKTTESKITATTTOVTSSTTQDTPFKITLTKTTTLAPKVTTKTKT 900  
Db 798 TAAPKMTKETATTTTEKTTESKITATTTOVTSSTTQDTPFKITLTKTTTLAPKVTTKTKT 857  
QY 901 ITTTEIMNKPEETAKPKDRATNSKATTPKPKQPTKAPKPKPTSTKKPKTMPRVRKPKTTP 960  
Db 858 ITTTEIMNKPEETAKPKDRATNSKATTPKPKQPTKAPKPKPTSTKKPKTMPRVRKPKTTP 917  
QY 961 PRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVNPKSEDAGGAEGETPHMLLRP 1020  
Db 918 PRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVNPKSEDAGGAEGETPHMLLRP 977  
QY 1021 HVFMPEVTPDMDYLPRVENQGIINPMLS 1049



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 115.337 Seconds  
(without alignments)  
5233.063 Million cell updates/sec

Title: SEQ1-D  
Perfect score: 5576  
Sequence: 1 MAWKTPPIVLLLLSVFVIQ.....DMDYLRVFNQGIINPMLS 1049

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.1

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Uniprot\_02: \*  
1: uniprot\_sprot: \*  
2: uniprot\_trembl: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5546.9	99.5	1404	2	Q92954
2	5523.9	99.1	1404	2	Q9BX49
3	3143	56.4	933	2	Q62M25
4	3143	56.4	933	2	BAD18580
5	2445.4	43.9	1054	2	Q9JW99
6	1346.8	24.2	5179	1	MUC2_HUMAN
7	1320.6	23.7	1225	2	Q9VR49
8	1314.2	23.6	1761	2	Q7KTF6
9	1314.2	23.6	1761	2	AAS64673
10	1243.6	22.3	3150	2	Q7PMD5
11	1169.4	21.0	3409	2	Q6SSE6
12	1169.4	21.0	3409	2	AAS07044
13	1128.7	20.2	1664	1	SLPI_CLOUTM
14	1113.7	20.0	1349	2	Q8WWQ4
15	1090.2	19.6	3432	2	Q8IR51
16	1090.2	19.6	3458	2	Q8IR52
17	1080	19.4	3889	2	Q6SSE8
18	1080	19.4	3889	2	AAS07042
19	1068	19.2	1795	2	Q76894
20	1051.3	18.9	1079	2	Q9N4S7
21	1034.6	18.6	9234	2	Q7KTP5
22	1034.6	18.6	9234	2	AAN10531
23	1020.5	18.3	2284	2	Q9VPG1
24	1020.4	18.3	972	2	Q7QKK7
25	1012.6	18.2	1607	2	Q8H6Q5
26	1012.6	18.2	1607	2	AAP74661
27	1002.8	18.0	1274	2	Q20007
28	993.7	17.8	1489	2	Q96449
29	990.9	17.8	10578	2	Q8ISF5
30	989.5	17.7	23015	2	Q8IQI8
31	989.5	17.7	23015	2	AAN10358

32 988.9 17.7 18519 2 Q8ISF6  
33 988.9 17.7 18534 2 Q8ISF7  
34 988.4 17.7 34350 2 Q8WZ42  
35 988 17.7 7962 2 Q10465  
36 978.7 17.6 5703 1 MUSB\_HUMAN  
37 975.9 17.5 2112 2 Q9VEI9  
38 974.8 17.5 2187 2 P70670  
39 964.4 17.3 926 2 Q9VVG2  
40 960.4 17.2 1720 2 Q8I486  
41 956.8 17.2 1480 2 Q9LIE8  
42 956.7 17.2 1458 2 Q757N5  
43 956.7 17.2 1458 2 Q757N5  
44 953.5 17.1 1121 2 AAS52662  
45 945.1 16.9 2042 2 Q7Z884  
Q76718 sus scrofa

## ALIGNMENTS

RESULT 1  
Q92954

ID Q92954 PRELIMINARY; PRT; 1404 AA.  
AC Q92954;  
DT 01-FEB-1997 (TREMblrel. 02, Created)  
DT 01-FEB-1997 (TREMblrel. 02, Last sequence update)  
DT 01-MAR-2004 (TREMblrel. 26, Last annotation update)  
DE Megakaryocyte stimulating factor.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,  
RA Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,  
RA Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,  
RA Bhatia S., Kriz R., Hewick R., Clark S.C.;  
RT "Purification, Biochemical Characterization, and Cloning of a Novel  
RT Megakaryocyte Stimulating Factor that has Megakaryocyte Colony  
RT Stimulating Activity.";  
RL Blood 78:279-279(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Merberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P.,  
RA Fitzgerald M., Scaltreto J., Kelleher K., Preissner K., Kriz R.,  
RA Jacobs K., Turner K.;  
RT "A Comparison of Vitronectin and Megakaryocyte Stimulating Factor.";  
RL (In) Preissner K.T., Rosenblatt S., Kost C., Wegerhoff J.,  
RL Mosher D.F. (eds.);  
RL BIOLOGY OF VITRONECTINS AND THEIR RECEPTORS., pp.45-52, Elsevier  
RL Science Publishers B.V. (1993).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,  
RA Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,  
RA Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,  
RA Bhatia S., Kriz R., Hewick R., Clark S.C.;  
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U70136; AAB09089.1; -;  
DR HSSP; P04004; 10C0.  
DR Genew; HGNC:9364; PRG4.  
DR GO; GO:0008283; P:cell proliferation; TAS.  
DR InterPro; IPR000585; Hemopexin.  
DR InterPro; IPR001212; Somatomedin\_B.  
DR Pfam; PF00045; Hemopexin; 2.  
DR PRINTS; PR00022; Somatomedin\_B; 2.  
DR SMART; SM00120; HX; 2.  
DR SMART; SM00120; SO; 2.  
DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN 1.  
DR PROSITE; PS00524; SOMATOMEDIN\_B; 2.  
SQ SEQUENCE 1404 AA; 151090 MW; AABD7AD19B35F4F6 CRC64;









```

Db 471 ATTEKTTESKIIATTTQVTSITTTQDTTPPKIITLLKTTTLAPKVTTTKKTIITTEIMNKP 530
QY 911 EETAKPKDRATNSKATTPKQKPTKAPKPKPTSTKKPKTMPRVRKPKTTTPRKMTSTWPE 970
Db 531 EETAKPKDRATNSKATTPKQKPTKAPKPKPTSTKKPKTMPRVRKPKTTTPRKMTSTWPE 590
QY 971 LNPTSRIAEAMLOTTTRENQTPNSKLVENVNPKSEDAGGAGETPHMLLRPHVFMPEVTPD 1030
Db 591 LNPTSRIAEAMLOTTTRENQTPNSKLVENVNPKSEDAGGAGETPHMLLRPHVFMPEVTPD 650
QY 1031 MDYLPRVNOGIIINPMLS 1049
Db 651 MDYLPRVNOGIIINPMLS 669

RESULT 4
BADI8580 PRELIMINARY; PRT; 933 AA.
AC BADI8580;
DT 12-MAY-2004 (TrEMBLrel. 27, Created)
DT 12-MAY-2004 (TrEMBLrel. 27, Last sequence update)
DT 12-MAY-2004 (TrEMBLrel. 27, Last annotation update)
DE CDNA FLJ16561 fis, Clone SYN04003981, moderately similar to Homo
DE sapiens proteoglycan 4, (megakaryocyte stimulating factor, articular
DE superficial zone protein, campodactyly, arthropathy, coxa vara,
DE pericarditis syndrome) (PRG4).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP TISSUE=Synovial membrane tissue;
RC Ota T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T.,
RA Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.
RA Yamaoka H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagatsuna M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagabari K.,
RA Masuo Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK131434; BADI8580.1; -.
SQ SEQUENCE 933 AA; 102512 MW; B883773C5BC21A23 CRC64;

Query Match 56.4%; Score 3143; DB 2; Length 933;
Best Local Similarity 56.2%; Pred. No. 3,7e-54;
Matches 618; Conservative 0; Mismatches 1; Indels 480; Gaps 2;

QY 1 MAWKTLPIYLLLSVFVIQVSSQELSCKGRCPESPFRGECDCDAQCKYDKCCPDYE 60
Db 1 MAWKTLPIYLLLSVFVIQVSSQELSCKGRCPESPFRGECDCDAQCKYDKCCPDYE 60
QY 61 SFCA-----EHSVS 70
Db 61 SFCAEVNTPSPSSKAPPPSGASQIKSTTKSPAPNKKTKKVIEEITEHSVS 120
QY 71 ENDESSSSSSSSSSSTWIKIKSSKNGAANRELQKLKVDKNKNRTKKKTPKPPVVDE 130
Db 121 ENQESSSSSSSSSSSTWIRIKSSKNGAANRELQKLKVDKNKNRTKKKTPKPPVVDE 180
QY 131 AGSLDNGDFKVTTPDSTTOHNKVSPTKITTAKPINRPSLPNSDTSKETSITWKE 190
Db 181 AGSLDNGDFKVTTPDSTTOHNKVSPTKITTAKPINRPSLPNSDTSKETSITWKE 240
QY 191 TTIVETKTTNTKQTSIDGKEKTTSAKETQSIETKSADLAPTSGVLAKPTPKAETTTKG 250
Db 241 TTIVETKTTNTKQTSIDGKEKTTSAKETQSIETKSADLAPTSGVLAKPTPKAETTTKG 300
QY 251 PALTTPKEPTTTPKEPASTTTPKEPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTP 310
Db 301 PALTTPKEPTTTPKEPASTTTPKEPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTP 360

```

```

QY 311 KEPAPTTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTTPKEPAPTTTPKEPPTP 370
Db 361 KEPAPTTTPKE----- 370
QY 371 TTPKEPAPTTTTPKEPAPTTTPKEPAPTTAPKKAPATTTPKEPAPTTTPKEPAPTTTKEPSPPTTPE 430
Db 371 ----- 370
QY 431 PAPTTPKSAPTTTPKEPAPTTTKSAPTTPKEPSPPTTTPKEPAPTTTPKEPAPTTTPKAPATTTP 490
Db 371 ----- 370
QY 491 KEPAPTTTPKEPAPTTTKKPAPTAPKEPAPTTPKETAPTTPKLIJTTTPEKLAPTTPEKPA 550
Db 371 ----- 370
QY 551 PTTPELAPTTTPEBPTTTPPEPAPTTTKAAANPTPKAPATTTPKEPAPTTTPKEPAPTTTP 610
Db 371 ----- 370
QY 611 KETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKBPTSTTSKAPATTTPKGTAPTT 670
Db 371 ----- 370
QY 671 PKEPAPTTTPKEPAPTTPKGTAPTTLKEPAPTTTPKKAPKELAPTTTTKGPTSTTSKAPATT 730
Db 371 ----- 370
QY 731 TPKETAPTTPPKAPATTTPKKAPATTTPETPPPTTSEVSTTTPKPTTIHKSPDESTPELS 790
Db 371 -----PAPTTPETPPPTTSEVSTTTPKPTTIHKSPDESTPELS 410
QY 791 ABPTPKALENSKPEPGVPTTKTAAATKPEMTTAKDKTTERDLRTTPETTTAAAPKVTKET 850
Db 411 ABPTPKALENSKPEPGVPTTKTAAATKPEMTTAKDKTTERDLRTTPETTTAAAPKVTKET 470
QY 851 ATTEKTTESKIIATTTQVTSITTTQDTTPPKIITLLKTTTLAPKVTTTKKTIITTEIMNKP 910
Db 471 ATTEKTTESKIIATTTQVTSITTTQDTTPPKIITLLKTTTLAPKVTTTKKTIITTEIMNKP 530
QY 911 EETAKPKDRATNSKATTPKQKPTKAPKPKPTSTKKPKTMPRVRKPKTTTPRKMTSTWPE 970
Db 531 EETAKPKDRATNSKATTPKQKPTKAPKPKPTSTKKPKTMPRVRKPKTTTPRKMTSTWPE 590
QY 971 LNPTSRIAEAMLOTTTRENQTPNSKLVENVNPKSEDAGGAGETPHMLLRPHVFMPEVTPD 1030
Db 591 LNPTSRIAEAMLOTTTRENQTPNSKLVENVNPKSEDAGGAGETPHMLLRPHVFMPEVTPD 650
QY 1031 MDYLPRVNOGIIINPMLS 1049
Db 651 MDYLPRVNOGIIINPMLS 669

RESULT 5
Q9JW99 PRELIMINARY; PRT; 1054 AA.
AC Q9JW99;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus.
GN Name=Pr94;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20573856; PubMed=11124536;
RA Ikegawa S., Sano M., Koshizuka Y., Nakamura Y.;
RT "Isolation, characterization and mapping of the mouse and human PRG4
RT (proteoglycan 4) genes.";

```







[2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22426065; PubMed=12537568;  
RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,  
Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,  
Paclet J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila  
melanogaster euchromatic genome sequence."  
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
[3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22426070; PubMed=12537573;  
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirskas R.,  
Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
Ashburner M., Celniker S.E.;  
RT "The transposable elements of the Drosophila melanogaster euchromatin:  
a genomics perspective."  
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
[4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22426069; PubMed=12537572;  
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,  
Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,  
Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
Lewis S.E.;  
RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
systematic review."  
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).  
[5]  
RP SEQUENCE FROM N.A.  
RG FLYBASE;  
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB003575; AAF50957.3; -.  
SQ SEQUENCE 1225 AA; 127735 MW; A99AF9D4404C79F4 CRC64;  
Query Match 23.7%; Score 1320.6; DB 2; Length 1225;  
Best Local Similarity 27.1%; Pred. No. 2.le-18;  
Matches 353; Conservative 119; Mismatches 416; Indels 414; Gaps 49;  
QY 10 LLLLLSVFVIQVS-----SQEL---SKGRCFSEFERGREGDCDAQCKYD 53  
DB 6 IFLTVSILLIQVKNRKYNDWSDMQGSPSEIIFGCGGDTIYYPDPVQPCDTPD-----58  
QY 54 KCPDYESCAEHSVSENQESSSSSSSSSTIWKIKSKNSAANRELQKLLK-----108  
DB 59 -----SNPTTPRQTKRKPCKSTRRTTKTKPRRX 89  
QY 109 -----VKONKNRKKKP-TPKPPVVDGAGSLDNGDFKVTTPDT-----147  
DB 90 TTKWTKRATKRTTKTRRRPTTKPT-----DTTDSPTITGAECTCSDRTTAS 140  
QY 148 ---STTOHNKVS---TSPKITAKPINRPSLPNPSDTSKETSIVNKEITVE-----194  
DB 141 STDSTDRVTWINTDWTPLCTDTPPCT-----CSEESSTAIPSSPCDITSVIP 190  
QY 195 ----TKETTTNKQTSDDGKEKT--TSAKETQSIKETSAKDLAPTSKVLAKPTPKAETT-247  
DB 191 TSPCTQETTTPTPCSTQGGTTPCTCAQT-----TTPRSTTTT 230  
QY 248 -TKGPAITTPKEPTTPKPSASTPKPEPTPTTIKSAFTTPKEPAPTTTKSAP--TTPKE 304  
DB 231 STSRPTTTTPRSTTTTTSRPTTTTPRSTTTTTPRSTTTTTPRSTTTTSCAPTTPRPS 290

QY 305 PAPTTPKEPAPTTPKEPAPTTPKBPAPTTTKSAPT--PKBPAPTTP-----KKP 352  
DB 291 TTTTTSRPTTTTPACITTTTSCPTRTPTSTTTTSRPTTTTPCTTTPTSTTCTTRP 350  
QY 353 APTTPKE-----PAPTTPKE-----PTPTTPKEPAPT--TKEPAPTTPKEPAPT 395  
DB 351 TTTTPRSTTTTCTSGPTTTTPRSTTTTCTSGPTTTTPRSTTTTSGPTTTTPRSTTTT 410  
QY 396 PKKPAPTTPKE-----PAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTKBEPA 447  
DB 411 TSGPTTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTTPRSTTT 470  
QY 448 PTTTKSAP--TTPKEPSPTTTKEPAPTTPKE-----PAPTTPKE-----PAPT 489  
DB 471 TTTTSGPTTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTT 530  
QY 490 PKE-----PAPTTPKEPAPTTPKAPAPKE-----PAPTTPKETAPTTPK 533  
DB 531 PRSTTTTSGPTTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTTPRSTTTTSGP 590  
QY 534 TPTTEKLAPTPKEPAPTTPPELAPTPEEPTTPPE-----PAPTTPKAAAPT 585  
DB 591 TTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTTPRSTTTT 650  
QY 586 PKEPAPTTPKE-----PAPTTPKE-----PAPTTPKETAP-----TTPKGT 621  
DB 651 TSGPTTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTTPRST 710  
QY 622 APTTLKEPAPTTPK-----PAPKELAPTTPKEPTSTT-----SDKPAPT 662  
DB 711 TTTTSGPTTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTT 770  
QY 663 PKGTAPTTPKEPAPTTPKE-----PAPTTPKGAPTTLKEPAPTTPK-----705  
DB 771 PRSTTTTSGPTTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTTSGPTTTTSGP 830  
QY 706 --PAPKELAPTTPKGTSTT-----SDKPAPTTPKETAPTTPKEPAPTTPK-----750  
DB 831 TTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTTPRSTTTT 890  
QY 751 ---PAPTTPET-----PPPTSEVSTPTTKETTHKSPDESTPELSAETPKALE 799  
DB 891 TSCPTTTTPRSTTTTCTSCPTTTTPRSTTTTSGPTTTTPRSTTTTCTSGPTTTT 944  
QY 800 NSPKPQVTTKTPAATKPEMTTAKDTERDLRTTBETTTAAKMTKETATTTEKITE 859  
DB 945 TTPRSTTTTSGPTTTTPRSTTTTTPRSTTTTTPRSTTTTTPRSTTTTTPRSTTT 999  
QY 860 SKITATTQVSTTTQDTPPKITTLKTTTLAPKVTTTKKITTITTEIMNKPEETA 919  
DB 1000 ---TTSRPTTTTPRSTT-----KTSCAPTTPRSTTTT-----TSRPTTT 1041  
QY 920 ATNSKATTPKQKPTKAPK-----KPTSTKKKPMRVRKPKTTPTRKMT-----967  
DB 1042 TPRSTTTTTSRPTTTTPRSTTTTPCTSRPTTTTPRSTTTTTPRSTTTTTPRSTTT 1101  
QY 968 MPENLNTSRAEAMLOTTR-----PNQT-----PNSKLVE 998  
DB 1102 TPSASPTR-----TTTTRPCPCHPQPPYQPPWSWYVNTYVFWVWQDN-----1149  
QY 999 VNPKSEDAAGAGETHEMLLRPHVFMPEVTPDMDYLPRVENQ 1040  
DB 1150 -----PVVPQWQMEGYENQ 1164

RESULT 8

Q7KTF6

ID Q7KTF6

AC Q7KTF6

DT 05-JUL-2004

DT 05-JUL-2004

DT 05-JUL-2004

PRELIMINARY;

PRT; 1761 AA.

(T-EMBLrel. 27, Created)

(T-EMBLrel. 27, Last sequence update)

(T-EMBLrel. 27, Last annotation update)





QY 442 --TTKEPA-----PTTKSA-----PTTPKEPSTTKPEA- 470  
DB 675 HETTAEPATKKTHTPTTKSTTLRITBEPTTKSSSTAKTTREPTTKKRETTERTTKEPTT 734  
QY 471 -----PTTPK-----EPA-----PTTPKKPAPTTP 490  
DB 735 RKTTHKTEBPTTKKTHPTTKKSTTLKPTKEPTTKSTTKTTREPTTKKKTERTT 794  
QY 491 KEPA-----PTTPKEPAPTTKKA--PTAPK-----EPA----- 518  
DB 795 KEPTTKKTHKTTBEPTNK-----TTTKKTHPTTKKSTTLKPTBEPTTKSTTKTT 850  
QY 519 -----PTTPKETA-----PTTPKLT-----PTT-PEK 540  
DB 851 REPTTKRETTERTTQEPSTKTTHTTAEPATKKTHTPTTKSTTLRITBEPTTKSS 910  
QY 541 LAPTPKEPAPTPEELAPTTPEPT-----PTTPE--EPA----- 574  
DB 911 TAKTTRE--PTTKRETTERTTKEPTTKKTHKTTBEPTTKKTHPTTKKSTTLKPT 967  
QY 575 -PTTPKAAPNTPKPAPTTPKEPAPTTPKEPA-----PTTPKETA-----P 615  
DB 968 EPTTKSTTKTRE--PTTKKKTERTTKEPTTKKTHKTTBEPTTKKKTERTTKEPTT 1025  
QY 616 TTPKGTGTA-----PTTPKEPAPTTPKKAPKELA-----PTTPKEPTSTSDKPA--- 659  
DB 1026 TTKKSTTLKPTBEPTTKSTTKTTREPTTKKVTERTTRETPTTKSTTKTTHTTAEPATKK 1085  
QY 660 -----PTTPKGTGTA-----PTTPKEPAPTTPKEPA-----PTTPKGTAPTTLKE 697  
DB 1086 TTHEPTTKSTTLRITBEPTTKSSSTAKTTREPTTKRETTERTTKEPTTKKTHKTTT 1145  
QY 698 PA-----PTTPK-----KPAPKELAPTTTKGPTSTSDKPAPTTPKAPTTPKEPA-- 744  
DB 1146 PTTKKTHPTTKKSTTLKPTB--PTTRK--TSTTKTREPTTKKKTERTTKEPTTR 1200  
QY 745 -----PTTPK-----KP-----APTTPPTTSEVST 768  
DB 1201 KTTTHKTTBEPTTKNTTKTTHTPTTKKSTTLKPTBEPTTKSTTKTTREPTTKRVT 1260  
QY 769 PTTPKEP-----TTPHKSPE-----STPELSAEPFKALENSPKPGVPTTKTAAATKP 818  
DB 1261 ERTRPTTKKTHKTTBEPTTKKTTTKTTHTPTTKK--STLKKEPTSEPTTKSTTKT 1319  
QY 819 EMTTAKDKTTERDLR-----TTPBTTTAAAP----- 844  
DB 1320 TREPTTKRVTERTTRETPTTKKTHKTTBEPTTKKTTTKTTHTPTTKKSTTLKPTB 1379  
QY 845 -----KMTKETAT--TTEKTESKITATTQVTSSTTQDTTPFKITTLKT----- 887  
DB 1380 TTKSTTKTTREPTTKRVTERTTRETPTTKT--THKTTBEPTTKKTTTKTTHTPTT 1437  
QY 888 ---TTLAP-----KVTTK-----KTIITTEIMNKP-----ESTAKPKDRATN 922  
DB 1438 KKSSTTLKPTBEPTTKKSTTKTTREPTTKKVTERTTRETPTTKKTHKTTBEPTTKKT 1497  
QY 923 SKATTPKQ-----KP-----TKAPKKPT----- 941  
DB 1498 TKKTHPTTKKSTTLKPTBEPTTKKSTTKATREPTTKKVTERTTRETPTTKKTHAHT 1557  
QY 942 -----STKKPKTMRVKPKKTTPTPRKWTSTMPELNPTSIAEAMLQ 983  
DB 1558 TESEPTTKKTTKTTHTPTTKKSTTL-----KPTBEPTTKSTTKTTREPTTKSTSVK 1613  
QY 984 TT-----TRENOTPNSKLVNPKSEDAAGAGET----- 1013  
DB 1614 TTADQTKRTTAEMSTTNQEPSTVETTTNSNQSNNTTSTTTTBEQHVHHHHHHVHKP 1673  
QY 1014 ----PMLLRPH----- 1021  
DB 1674 ADLGSILPLPLDPL 1733

QY 1022 VFMEPVT-PDMDYLPVPRVN 1039  
DB 1734 ISLSEISLNPPLPPLQPN 1752  
RESULT 9  
AAS64673 PRELIMINARY; PRT; 1761 AA.  
AC AAS64673;  
DT 01-APR-2004 (TRENBLrel. 27, Created)  
DT 01-APR-2004 (TRENBLrel. 27, Last sequence update)  
DT 01-APR-2004 (TRENBLrel. 27, Last annotation update)  
DE CG33300-PA (Fragment).  
GN CG33300.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
RN NCBI\_TaxID=7227;  
[1]  
SEQUENCE FROM N.A.  
RP MEDLINE=20196006; PubMed=10731132;  
RX Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.H., Blazek R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,  
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,  
RA Balow R.M., Basu P.V., Berman B.P., Brothier P., Brottier P.,  
RA Borkova D., Botchan M.R., Bouck J., Brocktein P., Brothier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
de la Fabia B., Delecher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,  
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pauley J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
RA Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195 (2000).  
[2]  
SEQUENCE FROM N.A.  
RP MEDLINE=22426065; PubMed=12537568;  
RX Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,  
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,  
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
RA Svirskas R., Tabor P.B., Wan K., Stapleton M., Sutton G.G., Venter C.,  
RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila  
melanogaster euchromatic genome sequence.";  
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079 (2002).  
[3]





Db	1558	TREPTTKTKTKTKTHBETTKKSTL-----KTEPTTRKTKSTYKTRPTRETTRETSVK	1613
Qy	984	TT-----TRRQTFNSKLVEYNPKSEDAGAGET-----	1013
Db	1614	TTADQTKRTTAEMSTTINQEPSVETTTNNSQSNNTTTESTTTEBQVHHHHHHIYHKP	1673
Qy	1014	-----PHMLLRPH-----	1021
Db	1674	ADLGSILPLDLPLPLFLPPLPLPELPLPLPTALPLPLPLPLPLPEVNLTA	1733
Qy	1022	VFMPEVT-PMDDYLPVRVN	1039
Db	1734	ISLPEISLPLPLPLELEN	1752
RESULT 10			
Q7PMD5			
ID	Q7PMD5	PRELIMINARY; PRT; 3150 AA.	
AC	Q7PMD5;		
DT	01-MAR-2004 (TReMBLrel. 26, Created)		
DT	01-MAR-2004 (TReMBLrel. 26, Last sequence update)		
DT	01-MAR-2004 (TReMBLrel. 26, Last annotation update)		
DE	ENSANGP00000004655 (Fragment).		
GN	Name=ENSANGG00000003651.		
OS	Anopheles gambiae str. FES.		
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;		
OC	Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.		
OX	NCBI_TaxID=180454;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=PEST;		
RA	Anopheles Genome Sequencing Consortium;		
RL	Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.		
CC	-!- CAUTION: The sequence shown here is derived from an		
CC	EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is		
CC	preliminary data.		
DR	EMBL; AAAB01008980; EAA13969.2; --		
DR	GO; GO:0016020; C:membrane; IEA.		
DR	GO; GO:0004872; F:receptor activity; IEA.		
DR	GO; GO:0008151; P:cell growth and/or maintenance; IEA.		
DR	InterPro: IPR006770; OGF receipt.		
DR	Pfam: PF04680; OGF_III; 80.		
FT	NON_TER	1	
FT	NON_TER	3150	
SQ	SEQUENCE	3150 AA; 322879 MW; 3C7B3D441CE8C839 CRC64;	
Query Match			
Best Local Similarity 22.3%; Score 1243.6; DB 2; Length 3150;			
Matches 402; Conservative 111; Mismatches 397; Indels 1114; Gaps 74;			
Qy	70	SENQESSSSSSSSSTTWIKSKSNANRELQKLVKONKNRTKKKTPKPPVD	129
Db	870	TESTDTTMSSASNPST-----TSGTTR-----TTPTPTPTDTTMS	907
Qy	130	EAGS-----GLDNGDFVVTPTDSTTQHN-----	153
Db	908	SASTPEPTPGTTRTTPRPTSTESTDTTMSSASTPEPTPGTTRTTPRPTSTESTD	967
Qy	154	-----KVSTSPKITTAKPINP-----RPSLP	174
Db	968	TMSSASTPEPTPGTTRTTPRPTSTESTDTTMSSASTPEPTPGTTRTTPRPTTT	1027
Qy	175	PNSDKESLTVNKETVETKETTNNKOTSDGKEKTSIA-----	216
Db	1028	ESTDTTMSSASTPEPTPGTTRTTPRPTSTESTDTTMSSASTPEPTPGTTRTTP	1086
Qy	217	--KETQSLEKTSAKDLAP---TSKVLAKPTPKAET-----TTKGPAL	253
Db	1087	RPTSTESTDTTMSSASTPEPTPGTTRTTPRPTSTESTDTTMSSASTPEPTPGTTR	1146
Qy	254	TTPKETPP-----TTPKEPAS-----	269
Db	1147	TTPTPTSTESTDTTMSSASTPEPTPGTTRTTPRPTSTESTDTTMSSASTPEPTSTP	1206
Qy	270	-----TTPKEP-----TPTTIKSAPTTP-----KEPA--	291
Db	1207	GTTTTPTRPTSTESTDTTMSSASTPEPTPGTTRTTPRPTSTESTDTTMSSASTPEPT	1266
Qy	292	PTTIKSAPTTP-----	306
Db	1267	PGTTRTTPRPTSTESTDTTMSSASTPEPTPGTTRTTPRPTSTESTDTTMSSASTPE	1326
Qy	307	PTTIKSAPTTPKEPAPT-----TTKEPA-PTTIKSAPTTP-----	341
Db	1327	PTTIKSAPTTPRPTSTESTDTTMSSASTPEPTPGTTRTTPRPTSTESTDTTMSSA	1386
Qy	342	KEPAP-----TTPKKPAPT-----TPKEPAPTTPKEPTTTTPKEPAPT-----	379
Db	1387	STPEPTPGTTRTTPRPTSTESTDTTMSSASTPEPTPGTTRTTPRPTSTESTDTT	1446
Qy	380	-----TKEPAP-----TTPKEPAPT-----APKKPAP-----TTPKEPAP-----	409
Db	1447	MSSASTPEPTPGTTRTTPRPTSTESTDTTMSSASTPEPTPGTTRTTPRPTSTES	1506
Qy	410	-----TTPKEPAPTTPKEPAPTTPKEPA-----PTTT	436
Db	1507	TDITMSSAYTPDPGIPGTTRTTPRPTSTESTDTTMSSASTPEPTPGTTRTTPRPTST	1565
Qy	437	KSAPT-----TKEPAP-----PTTIKSAPTTP-----KEPSTTTKEPAPTTPKE	476
Db	1566	ESTDTTMSSASTPEPTPGTTRTTPRPTSTESTDTTMSSASTPEPTPGTTRTTPR	1625
Qy	477	PAP-----TTPKKPAP-----TTPK	491
Db	1626	PTSTESTDTTMSSASTPEPTPGTTRTTPRPTSTESTDTTMSSASTPEPTPGTTRTTP	1685
Qy	492	EPAPT-----TPKEPAPTTPKEPA-----PTAP-----KEPAPT	520
Db	1686	RPTPTDSTMSSMSSESTPEPTPGTTRTTPRPTSTESTDTTMSSASTPEPTST	1745
Qy	521	TPKETAPTTPKLTPT-----TPE-----KLAPT-----T	545
Db	1746	TTPTGTTTPTPTSTESTDTTMSSASTPEPTPGTTRTTPRPTSTESTDTTMSSAST	1805
Qy	546	PE-----KPAPTTPBELAPTTPPEPTPT-----	568
Db	1806	PEPTTPGTTRTTPRPTSTESTDTTMSSASTPEPTPGTTRTTPRPTSTESTDTTMS	1865
Qy	569	-----TPEE-----PAPTTPKAAAPNPKPAP-----	591
Db	1866	SASTPEPTPGTTRTTPRPTSTESTDTTMSSASTPEPTPGTTRTTPRPTSTESTD	1925
Qy	592	-----TTPKEPAP-----TTPKEPAPTTPK	611
Db	1926	TMSSASTPEPTPGTTRTTPRPTSTESTDTTMSSASTPEPTPGTTRTTPRPTSTEST	1985
Qy	612	ETA-----PTTPKGTAPTLLKEPAP-----TTPKKPAP	639
Db	1986	DTTMSSASTPEPTPGTTRTTPRPTSTESTDTTMSSAYTPPEPTPGTTRTTPRPT	2045
Qy	640	KELAPT-----TKETST-----TSDKPAPTTPKGTAPTTPKEP-	674
Db	2046	TESTDTTMSSASTPEPTPGTTRTTPRPTSTESTDTTMSSASTPEPTPGTTRTTPRPT	2105
Qy	675	-----APTTP-----KEPAPTTPKGTAPT	694
Db	2106	STESTDTTMSSASTPEPTPGTTRTTPRPTSTESTDTTMSSASTPEPTPGTTRTTP	2165
Qy	695	LKEPAP-----TTPKKPAPKELA-----PTTIKGTSTT	723
Db	2166	PTPTPTDSTMSSMSASPTPGTTRTTPRPTSTESTDTTMSSASTPEPTPGTTRTTP	2225
Qy	724	SDKPAPT-----TPKE-----TAPTTPKAPAP-----	745
Db	2226	PTPTPTDSTMSSMSASPTPEPTPGTTRTTPRPTSTESTDTTMSSMSASTPEPTSTP	2285

QY	746	----	TTPKKPAP	----	TTPE	----	757
Db	2286	GTTRTTPRTPTDSTMS	SSSSASTPESSTTPE	LLQDQHLPIQCLVCRQ	PLLRSH	P	2345
QY	758	----	----	----	----	----	----
Db	2346	RHLVQELLQDQHLPI	POCRQFLRSHRLVQ	PELLQDQHLPLSTDT	TMSSASTPE	S	2405
QY	763	T	-----	SEVSTPTTTKEPTL	HKSP	-----	-----
Db	2406	ITPGTTRTTPRTPTD	TMSSASTPESTTPT	GTTRTTPRTPTST	TESTDTMSSAS	TPE	2465
QY	790	SNAP	-----	-----	TPKALENSKPEP	-----	-----
Db	2466	STTPGTTRTTPRTPT	STESTDTMSSASTPE	SPSTTPTGTRTTP	PRPTSTDTMTAS	-----	-----
QY	816	TKPEMTTAKDKTTER	DLRTPTTETTAAPKM	TKETATT	-----	-----	-----
Db	2526	T-PEPSIT	-PDTTRTTPRTPTD	TMSSASTPESTTPT	GTTRTTPRTPTD	TMSSAS	2583
QY	854	-----	TEKTESKITATT	TOVSTTODTTPK	TLTKTTLAP	-----	-----
Db	2584	TPEPSTTGTTRTTP	RSTSDTMMSSMSAS	TPESTTGTTRTTP	RPTPTDSTMS	2643	-----
QY	893	-----	KVTTKKIITTEIMN	KEETAKPKDRATN	-----	-----	-----
Db	2644	SMSSASATPGTTRT	TPRTPTSTESTDTM	SSASTPESTTPTG	TTRTTPRTPTD	STMS	2703
QY	923	--SKATPKQ--	-----	KPT	-----	KAPKKPTSK	-----
Db	2704	SMSSASTPESTTGT	TTRTTPRTPTD	STMSMSASTTPTG	TTRTTPRTPTST	2763	-----
QY	945	-----	KPKTMP	---RVRKPTTPER	KWTSM	-----	-----
Db	2764	TMSSASTPESTTGT	TTRTTPRTPTD	TMSSMSASTPESTT	PTGTRTTPRTPT	2817	-----
QY	986	TRPNOTNSKLVEVP	KSEDAGGAEGETPH	MLLRPHVFMDEVTP	1029	-----	-----
Db	2818	TRPTDSTMSMSAS	-----	TPE	-----	PSITP	2843
RESULT 11	Q6SSE6	PRELIMINARY;	PRT;	3409	AA.		
ID	Q6SSE6	AC	Q6SSE6;				
DT	05-JUL-2004	(T	REMLrel. 27, Created)				
DT	05-JUL-2004	(T	REMLrel. 27, Last sequence update)				
DE	plus agglutinin.						
GN	Names=SAGI;						
OC	Chlamydomonas reinhardtii.						
OC	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;						
OC	Chlamydomonadaceae; Chlamydomonas.						
RN	NCBI_TaxID=3055;						
OX	[1]						
RP	SEQUENCE FROM N.A.						
RA	Ferris P.J., Waffenschmidt S., Umen J.G., Ishida K., Kubo T., Lau J.,						
RL	Goodenough U.W.;						
DR	EMBL: AY450930; AAS07044.1; --						
DR	InterPro; IPR02951; Atrophin.						
DR	PRINTS; IPR003882; Pistil_extensin.						
DR	PRINTS; PRO1222; ATROPHIN.						
DR	PRINTS; PRO1218; PSTLEXTENSIN.						
SQ	SEQUENCE 3409 AA; 336045 MW; 7FE87633EDD6631F CRC64;						
Query Match	21.0%; Score 1169.4; DB 2; Length 3409;						
Best Local Similarity	15.5%; Pred. No. 6.7e-15;						
Matches 345; Conservative	149; Mismatches 403; Indels 1336; Gaps 63;						
QY	1	MAWKTLPI	-----	YLLLSLV	-----	-----	16



Db	621	-----PPSP	624
QY	236	VLAKPTPKAETTTKGPAITTPKEPTTTTPKEPAST-----TPKEPTTTIKSAPTTPK	288
Db	625	APSPAPPS-----PAPSPQPPSPVPQPPSPVPSPKPPSPAPPSVPVPPSPAPPSPA	678
QY	289	EPAPTTKSAPTTKKEPAPTTPKAPAPTTPK-----EPAPTTPKAPAPTTPKAPAPTTPK	344
Db	679	PPSPAPPNAPPSAPPLPPSPDEPPSPAPPSPEPPSPAPPSAPPSAPPSAPPSAPPS	738
QY	345	APTTPKAPAPT-----TPKEPAPTTPKAPAPTTPKAPAPTTPKAPAPTTPK	390
Db	739	APPSPEPPSPAPPSPEPPSPAPPSAPPSAPPSAPPSAPPSAPPSAPPSAPPS	798
QY	391	PAPT-----APKPAAPTTPKAPAPTTPKAPAPT-----	418
Db	799	PSAPPSAPPSPEPPSPAPPSPEPPSPAPPSPEPPSPAPPSPEPPSPAPPSPEPPSP	858
QY	419	TTKEPSTTPKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTPK	477
Db	859	PAPPSAPPSPEPPSPAPPSPEPPSPAPPSPEPPSPAPPSPEPPSPAPPSPEPPSP	917
QY	478	AP-----TPKKEAPTTPKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTPK	515
Db	918	SPERSPEPPSPAPPSPEPPSPAPPSPEPPSPAPPSPEPPSPAPPSPEPPSPAPPS	977
QY	516	-----EPAPTTPKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTPK	556
Db	978	PPSPDPSPAPSPDPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP	1036
QY	557	LAPT-----TPEPTTPTPKEPAPTTPKAAAPN-----TPKEPAPTTPK	595
Db	1037	TPPTSP	1096
QY	596	EPAPTTPK-----EPAPTTPKAPTTPKGTAPTL-----	626
Db	1097	PPSPAPSP	1156
QY	627	-----KEPAPTTPKAPKAPKELAPT-----TKEPT	651
Db	1157	SPEVPPTTPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP	1216
QY	652	STTSKAPAPTTPKAPTTPKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTPK	711
Db	1217	SPEPPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP	1271
QY	712	APT-----TTKGTSTTSKAPAPTTPKAPTTPKAPAPTTPKAPAPTTPKAPTTPKAPTTPK	769
Db	1272	SPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP	1328
QY	770	TTTKEPTTIHKSPDESTPELSAETPKALE-----NSPKKEGV- 807	
Db	1329	AAALPPLPPSPAPPLVPVPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP	1388
QY	808	-----PTTKTP-----AATKPE-----	819
Db	1389	PPPTPTTPPLAPLPDCTLLAQALLSIPDAANSVFVVSAGLTPSVAPSTPELLASFCT	1448
QY	820	-----MTTA-----	824
Db	1449	VCSCQTATATSLVGGSSRRNNTNGSSGNYYSNGGDAAIQIPAGNSTRDGGSSSGS	1508
QY	825	-----KDKTTERDL-----RTTP-----	837
Db	1509	GSSSWGPGTETAEMAVDAVDQGTQQLSIGGVYTRTVVDRTPPSVSGNVTLSANRI	1568
QY	838	-----ETTTAAPTWT-----	847
Db	1569	KOEPSAVGEASLNALGSKQAMLLTISFSEVPAPFAPASLIIVTGALVAEWAADKMTFY	1628
QY	848	-----KETATTTEKTTESKIT-----	863
Db	1629	VLAAMLPAELVATAAGSSSSGSTRSGNGNGTAAAAAAPPAGTTGRRALQOQAAA	1688

RESULT 13  
SLP1\_CLOTH STANDARD; PRT; 1664 AA.  
AC Q06852;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, last sequence update)  
DT 05-JUN-2004 (Rel. 44, last annotation update)  
DE Cell surface glycoprotein 1 precursor (Outer layer protein B) (S-layer protein 1).  
GN Name=olpB;  
OS Clostridium thermocellum.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1515;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCIB 10582;  
RX MEDLINE=93209931; PubMed=8456832;  
RA Fujino T., Beguin P., Aubert J.-P.;  
RT "Organization of a Clostridium thermocellum gene cluster encoding the cellulosomal scaffolding protein Cipa and a protein possibly involved in attachment of the cellulosome to the cell surface."  
RL J. Bacteriol. 175:1891-1899(1993).  
CC -!- SUBUNIT: Assembled into mono-layered crystalline arrays.  
CC -!- SUBCELLULAR LOCATION: Cell wall.  
CC -!- SIMILARITY: Contains 4 S-layer homology (SLH) domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).











```
QY 692 -PTTLKEPAPTPK-----KPA-----PRELAPT-----TTKGPT 720
Db |||||
QY 892 KPTTLKPTTEGTAKPTTLKPDGTAKPTTLKPTTEGTSAKPTTLKPDGTAKSTTLKPT 951
Db |||||
QY 721 STTSDDKA-----PTTPKETAPTTKEPAPTPK-----
Db |||||
QY 952 EGTSAKPTTLKPTTEGTAKPTTLKPTTEGTAK-----PTTLKPTTEGTSAKPTTLKPTTEGTTA 1008
Db |||||
QY 750 -----KPA-----PTT-----PETPPPTTSEVSTPTTK-----
Db |||||
QY 1009 KPTTLKPTTEGTSAPPTTLKPTTEGTSAPPTTLKPTDGTAKPTTLKPTTEGTSAPPTTLKPT 1068
Db |||||
QY 774 -----EPTTIHKSDESPELSAPPT-----KALENSKPEPGVPTTKPAATKPEMTT 822
Db |||||
QY 1069 EGTTAKPTTL--KPTTEGT--SAOPTTLKPTTEGTSAPPTTLKPTTEGTTAKPTTLKPTTEGT 1123
Db |||||
QY 823 TAKDKTTERDLRITPTTT-----AAPKMTKETATTTKTESKITATTT-----
Db |||||
QY 1124 SAKPTTLKPTTEGTTAKPTTLKPTTEGTSAPPTTLKPTTEGTTAKPTTLKPTTEGTTAKPTTLK 1183
Db |||||
QY 868 -----QVTSITTTQDIT-----PFKITTLKTTTAP-----K 893
Db |||||
QY 1184 PTEGTSAPPTTLKPTTEGTTAKPTTLKPTTEGTSAPPTTLKPTTEGTTAKPTTLKPTTEGTSAP 1243
Db |||||
QY 894 VTTTKKTIITTEIMNKPEETAKPD-----RATNSKATTPKPKOKTKAPKKPT 941
Db |||||
QY 1244 PTTLKPTTEGTTA--KP-TTLKPTDGTAKPTTLKPTTEGTSAPPTTLKPTTEGTTA--KPT 1297
Db |||||
QY 942 -----STKKPKTM-----PRVRKP-----KITPTPRKMT 965
Db |||||
QY 1298 TLKPTTEGTTAKPTTLKPTDGTAKPTTLKPTTEGTSAPPTTLKPTTEGTTAKPTTLKPTTEGT 1357
Db |||||
QY 966 STMP-ELNPTSRIAEAMLOTTTRPNOTPNPSKLVNPNKSEDAGGARGETPHMLLRPHVEM 1024
Db |||||
QY 1358 SAKPTTLKPTD--GTTAKPTTLKPTTEGTSAPPTTLKPT-----TEGTT-----AKPTTLK 1404
Db |||||
QY 1025 PEVTFDMDYLPRV--PNQGIINP 1046
Db |||||
QY 1405 P--TEGTSAPPTTLKPTTEGTTAKP 1426
```

Search completed: October 13, 2004, 12:18:04  
Job time : 142.337 secs

**This Page Blank (uspio)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 21.3671 Seconds  
(without alignments)  
4723.689 Million cell updates/sec

Title: SEQ1-D  
Perfect score: 5576  
Sequence: 1 MAWKTLPIVILLLLSVFVIQ.....DMDYLPVFNQGIINPMLS 1049

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.1

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 79:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1283.1	23.0	3020	2	A43932	mucin 2 precursor,
2	1128.7	20.2	1664	2	T18262	S-layer protein -
3	1002.8	18.0	1274	2	T16251	hypothetical prote
4	993.7	17.8	1489	2	T31108	cyst germination s
5	988	17.7	7962	2	T38346	elastic titin - hu
6	975.4	17.5	3570	2	T45025	mucin MUC5B, trach
7	974.8	17.5	2187	2	T30826	nascent polypeptid
8	918	16.5	1367	1	S48478	glucan 1,4-alpha-g
9	871.4	15.6	1188	2	S49915	extensin-like prot
10	861.6	15.5	1229	2	T25697	hypothetical prote
11	860.6	15.4	3507	2	T34513	hypothetical prote
12	857.5	15.4	6642	2	T29757	protein UNC-89 - C
13	835.4	15.0	2897	2	E48666	cell proliferation
14	834.9	15.0	3256	2	A48666	cell proliferation
15	826.9	14.8	5762	2	A41819	proline-rich pepti
16	822.8	14.8	1151	2	T18535	high molecular mas
17	818.2	14.7	1344	1	A35175	mucin 1 precursor,
18	768.9	13.8	2232	2	T34434	hypothetical prote
19	768.2	13.8	4135	2	T42629	tenascin-X - bovin
20	762.9	13.6	1832	2	T31113	mucin-like glycopr
21	756.1	13.6	2142	2	B35098	MHC class III hist
22	746.2	13.4	3942	2	T42730	Bassoon protein -
23	744.5	13.4	990	2	I51618	nucleolar phosphop
24	740.2	13.3	5262	2	T03454	ALR protein - huma
25	739.5	13.2	2225	2	T26063	hypothetical prote
26	733.9	13.2	761	2	C84672	hypothetical prote
27	733.5	13.2	924	2	S27923	Gene Lf3 protein -
28	733.4	13.2	4548	1	S00657	apoptosis(a) (EC
29	731.3	13.1	971	2	T19431	hypothetical prote

## ALIGNMENTS

## RESULT 1

A43932  
mucin 2 precursor, intestinal - human (fragments)  
N: Alternate names: mucin SMUC-41  
C: Species: Homo sapiens (man)  
C: Date: 10-Mar-1993 #sequence revision 12-Apr-1996 #text change 09-Jul-2004  
C: Accession: A49963; A45106; B45106; A43932; B35532; A61257; P00328; P00329  
J: Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.  
J: Biol. Chem. 269, 2440-2446, 1994  
A: Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the  
A: Reference number: A49963; MUID:94132002; PMID:8300571  
A: Accession: A49963  
A: Molecule type: mRNA  
A: Residues: 1-639 <GUI>  
R: Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Roche, E.M.; Lagace, R.E.; Kim, Y.S.  
J: Biol. Chem. 267, 21375-21383, 1992  
A: Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both upstr  
A: Reference number: A45106; MUID:93016075; PMID:1400449  
A: Accession: A45106  
A: Status: not compared with conceptual translation  
A: Molecule type: mRNA  
A: Residues: 626-1895 <GUI>  
A: Cross-references: GB:M94131; NID:G186395; PIDN:AAA59163.1; PID:G186396  
A: Note: sequence extracted from NCBI backbone (NCBIP:116706)  
A: Accession: B45106  
A: Status: not compared with conceptual translation  
A: Molecule type: mRNA  
A: Residues: 2037-3020 <GUI>  
A: Cross-references: GB:M94132; NID:G186397; PIDN:AAA59164.1; PID:G186398  
A: Experimental source: colon  
R: Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen, G.M  
J: Clin. Invest. 88, 1005-1013, 1991  
A: Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polymorp  
A: Reference number: A43932; MUID:91358717; PMID:1885763  
A: Accession: A43932  
A: Molecule type: DNA  
A: Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>  
A: Cross-references: GB:M74027; NID:G188863; PIDN:AAA59875.1; PID:G188864  
A: Note: sequence inconsistent with the nucleotide translation  
A: Note: sequence extracted from NCBI backbone (NCBIP:55749, NCBIP:55750)  
R: Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lampert, D.T.A.; Kim, Y.S.  
J: Biol. Chem. 264, 6480-6487, 1989  
A: Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and eviden  
A: Reference number: A33532; MUID:89197956; PMID:2703501  
A: Accession: B33532  
A: Molecule type: mRNA  
A: Residues: 1916-2193 <GUI>  
A: Cross-references: GB:M22405; NID:G188873; PIDN:AAA36334.1; PID:G188874  
A: Experimental source: intestine  
R: Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.

MHC class III hist  
MHC class III hist  
microtubule-associ  
Bassoon protein -  
hypothetical prote  
hypothetical prote  
hypothetical prote  
ascites sioglyco  
probable tenascin  
ALR protein - huma  
hypothetical prote  
All-1 protein +GTE  
cell proliferation  
mucin, tracheobron  
versican precursor  
hypothetical prote

seq1-d.rpr

Wed Oct 13 12:38:23 2004

J. Clin. Invest. 87, 77-82, 1991  
A>Title: Human bronchus and intestine express the same mucin gene.  
A/Reference number: A61257; MUID:91086481; PMID:1985113  
A/Accession: A61257  
A/Status: not compared with conceptual translation  
A/Molecule type: mRNA  
A/Residues: 1925-1948, 'TTS', 1952-1954 <JAN>  
A/Experimental source: bronchus  
A/Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCool, D.; Wang, D.; Jones, C.; Forstner, R.; Biochem. Biophys. Res. Commun. 183, 821-828, 1992  
A/Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-terminus  
A/Reference number: P00328; MUID:92198477; PMID:1550588  
A/Accession: P00328  
A/Molecule type: mRNA  
A/Residues: 2328-2468 <XUG>  
A/Cross-references: GB:M86523  
A/Experimental source: small intestine  
A/Accession: P00329  
A/Molecule type: protein  
A/Residues: 2328-2342, 'K', 2344-2354 <XUG1>  
C/Genetics:  
A/Genes: GDB:MUC2  
A/Cross-references: GDB:120203; OMIM:158370  
A/Map position: lip15.5-lip15.5  
C/Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von Willebrand factor type C repeat homology  
C/Keywords: glycoprotein; intestine; tandem repeat  
F:2746-2834/Domain: von Willebrand factor type C repeat homology <WVC>  
Query Match 23.0%; Score 1283.1; DB 2; Length 3020;  
Best Local Similarity 21.1%; Pred. No. 2.7e-18;  
Matches 393; Conservative 98; Mismatches 482; Indels 889; Gaps 57;  
QY 3 WK-----TLPIVLLLSVFVIQVSSOELCKRCFSF-----E 38  
DB 649 WREHVNKDVGCPSQVFLNYLTTCQTCRSLSEADSHLEGFAPVDGCGCPDHTFLDE 708  
QY 39 RGR-----ECDC-----DAQCKYDKC----- 55  
DB 709 KGRVPLAKSCVHRGLYLEAGDVVVRQBERCVCRDRLHCRQLRIGOSCTAPKIHMC 768  
QY 56 -----CPD----- 58  
DB 769 SNLTALATSKPRALSCQTLAGYHYTECVSGCPCDGLMDGRCVVEKFCVHNNDL 828  
QY 59 ----- 60  
DB 829 YSSGAKIKVDCNTCKRGRWVCTQAVCHCTSIYSGHYITFDGKYVDFDGHCSYAVQ 888  
QY 61 SFCAEHS-----VSNQSSSSSSSSSSSIWIKSSKNSANRELQKLVKDNKK 114  
DB 889 DYCGNSSLSGFSIITENVPGTGTGVCASKAIFMGRT-----ELKLEDKHR 936  
QY 115 -----NRT-----KKKPPKPPVVDAGSGL 135  
DB 937 VTIQDEGHVAVYTRVGVQLVVESTGLIIVDXKETTIFIKLAPSYKGV-----CGL 991  
QY 136 -----DNGDFKV-----TPDSTTQ----- 151  
DB 992 CGNFDHRSNNDFTTRDHMVVSSELDFGNSWKEAPTCPDVSTNPPCSLPHRRSWARKQC 1051  
QY 152 -----HNKVSTSP----- 159  
DB 1052 SILKSSVFSICHSKVDPPFYACVHDSCTDGDCEFCFASVASYAQBCTKGACVFW 1111  
QY 160 -----KITAKPIN----- 168  
DB 1112 RTPDLCPICFYDYNPPHCEWHEPCGRNSETCTINGIHSNISVYLEGCPKPCXDR 1171  
QY 169 -----PRSLP----- 174  
DB 1172 PIYEEDLKKVTDKCGCVEDTHYPGASVPTETCKSCVCTNSQVVCPRPEGIKLNQ 1231  
QY 175 -----PNSDTSKE-----TSLTVNKET-----TVETKETTITNKQSTSD 208

DB 1232 TDGAFCYWEICGNGTGVKXFNICSIITRPSLTITTTTLTPTTSFTTTTTPTTS 1291  
QY 209 GKEKTTISAK-----ETQSIKTSAD----- 229  
DB 1292 STVLSTTPKLCCLWSWDINEDHPSSGSDGDRPFDDGVCAGPEDIKRSVKDPHLSLEQH 1351  
QY 230 -----LAPTSKVLAKPTPKAE 245  
DB 1352 GQKVCQDVSVGFIKCNEDQFGNGPFLCYDYKIRVNCWMDKICITTPSPPTTTPSPPT 1411  
QY 246 TTTKGPALTTPKETPTTPKEPASTTPKEPPTTIKSAP-TTPKEPAPTTTKSAPTTPKE 304  
DB 1412 TTTTLPPTTTPSPPTTTTTPPTTTPSPPTTTTTPPTTTPSPPTTTPPTTTPS 1471  
QY 305 PAPTTTKEPAPTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTPK 355  
DB 1472 P-PITTPSPPTTTPSPPT-TTTTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPS 1529  
QY 356 TPKEPAPTTTPKEPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 415  
DB 1530 TTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPS 1589  
QY 416 APTTTKEPSPPTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEP 475  
DB 1590 TTTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPS 1641  
QY 476 EPAPTTTPKAP-TTPKEPAPTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTPK 534  
DB 1642 SPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPS 1695  
QY 535 PTTPEKLAAPTTPKEPAPTTPELAPTTPPEPPTTPPEPAPTTTPKAAAPNTPKPA 590  
DB 1696 WTPS-PITTPSSPITTTTPSSSITTP-SPPPTTMTTPSPTT-----TPSPPTTMT 1745  
QY 591 PTTTPKEPAPTTTPKEPAPTT-----KETAPTT-----KG----- 620  
DB 1746 TLPTTTPSSPLTTLPLPSITPTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPS 1805  
QY 621 ----- 620  
DB 1806 TELIGVCGGWAANISCRATWYDPVPIGQLGQTVVCDVSVGLICKNEDQKPGVPMAP 1865  
QY 621 -----TAPTTKEPAPTTTPKAPKELAPTT-TKETSTSTSDK-----AP 660  
DB 1866 CLNVEINVQCECVTQPTTM-----TTTTTENPTPTTTPPTTTPPTTTPPTTTPS 1922  
QY 661 T-TPKGTAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKAPKELAPTT 719  
DB 1923 TPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPS 1981  
QY 720 TST-TSDKAPTTTPKETAAPTTPKEPAPTTTPK-----APTTPPTTTPSEVSTPTT 771  
DB 1982 TPPTSTKSTVTPITTTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPS 2041  
QY 772 TK-----EPTTIHKSDESPKALNSPKPVGPTTKTAAKPEMTTAKD 826  
DB 2042 TSTAPIAELTSSNPPSSSTPQTSRSTSPSTTLLSTLTPALEMTSTAPSTPTAPT 2101  
QY 827 KTTTERTLTPPTTTPKAPKKTATATTEKTSKITATTQTSTTQDTTPPKIT 884  
DB 2102 TTSGGHTLSPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPS 2159  
QY 885 LKTTTTLAPKVTTPKTTTITTEIMNKPEE-----TA 914  
DB 2160 IIRITGLRPFSSVLICCVLNDTYVAFGEVYNGYGDTCYFVNCSLSTLEFYNWSCPS 2219  
QY 915 KPKDRATNSKATTPKAPKPTSTKPKTMPVRKPKTTPTRKMTSTWPELNPT 974  
DB 2220 TPTSTPTPSK-STPTPSKPSSTPSKPTPGTKPEPCDFDPPR-----QENET 2265  
QY 975 SRIAEAMLOTTTRNQITNSKLVEVNP-----KSEDAGGA----- 1009

Db 2366 WMLCCEM-ATCKYNNITVEIUVKCEBPPMPTCSNGLQVRVEDPDGCGHWECDYCTG 2324  
Qy 1010 -----EGETHMLLRPHVFPVTPDMD----- 1032  
Db 2325 WGDPHYVTFDGLYYSQNCYVLVE-----EISPSVDNGVYDYNHCDNDKVS CPR 2378  
Qy 1033 -----YLPVFNQGI 1043  
Db 2379 TLIVRHBTQEVLIKTVMHMQVQVNRQAVLPYKKYGLVYQSGINYYVDIPELGVL 2438  
Qy 1044 IN 1045  
Db 2439 VS 2440  
  
RESULT 2  
T18262  
S-layer protein - Clostridium thermocellum  
C;Species: Clostridium thermocellum  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T18262  
R;Fujino, T.; Beguin, P.; Aubert, J.P.  
J. Bacteriol. 175, 1891-1899, 1993  
A;Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulose  
e.  
A;Reference number: Z18847; MUID:93209931; PMID:8458832  
A;Accession: T18262  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1664 <FUG>  
A;Cross-references: UNIPROT:Q06852; EMBL:X67506; NID:g296879; PID:g296881; PIDN:CAA47841  
  
Query Match 20.2%; Score 1128.7; DB 2; Length 1664;  
Best Local Similarity 21.6%; Pred. No. 1.4e-15;  
Matches 366; Conservative 126; Mismatches 418; Indels 783; Gaps 67;  
  
Qy 4 KTLPIVLLLSVFFV-----IQVSSQ 26  
Db 6 KVLISLLTLLIISTTSVNMSPAEATPSIEMVLDKTEVHGVITATIKVNNIRKLAGYQ 65  
Qy 27 LSCK-----GRCFSEFREGRECDCAQCKYDKCP----- 57  
Db 66 LNIKFDPEVLQVDPATGEEFT-----DKSNPVNRVLLTNSKYGPTPV 108  
Qy 58 -----DYESFCAERHS-----VSENOES-----SSSS 80  
Db 109 AGNDIKSGIINPATGVNNLTAYKSSGIDEHTGIGIEGKVLKQNTSIRFEDTLSPGA 168  
Qy 81 SSSSSSTIW-----KIKSKNSAANRELQK-KLKVKD-----N 112  
Db 169 ISGTSLFDWDAETITGVEVIQPDLIWBAEPLKDSVALELDKTKVKVGDIITAIKIKEN 228  
Qy 113 KKN-----RUKKKT-----PKPPVDEAGSG 134  
Db 229 MKNFAGYQINIKYDPTMLEALETGSIAKRTWPVGTGLQSDNYGKTTAVANDVGAG 288  
Qy 135 LON-----GDFKVTTPD----- 147  
Db 289 IINFARAYSNLTKYRETGVABETGIIGKGRVLKAGSTAIRFEDTAMPGAIGETMYFD 348  
Qy 148 -----SITQHNK-----STSPKITTAKPINRPSLP-----PNS--- 177  
Db 349 WYGENIKGYSVQVQGEIVAEGEERPEEPTVPTETVDTPTVTEEPVPSLPDSVI 408  
Qy 178 ---DTSK-----ETSLTVNKKETTVEK 196  
Db 409 MELDKTKVKVGDIITAIKIKENKNFAGYQINIKYDPTMLEALETGSIAKRTWPVGT 468  
Qy 197 ETTTNNKQSTDGKRTTS-----AKETQSIEKTSKD 229  
Db 469 GTV-----LQSDNYGKTTAVANDVGAGIINPAEAYSNLTKYRTGVABETGIIGKGRV 523  
Qy 230 LAPTSKVLAKPTPKAETTT-----KGFALTTP-----KEP 259

Db 524 LKAGSTAI-----RFEDTTAMPGAIEGTYMPDWYGENIKGYSVQPGSIVAEGEPTPEP 578  
Qy 260 TPT-TPKEPASTTKPEPTPT----- 278  
Db 579 VPTETVDPPTPTVTEEPVPSLPDSYIMELDKTKVKEGDVIIATIRVNNIKNLAGYQIG 638  
Qy 279 ----- 278  
Db 639 IKYDPKLEAFNIETGDPIDEGTWPVAVGGTILKNRDYLPCTGVAINNVSKGLINFAAYVY 698  
Qy 279 -----TIKSAPTPK----- 288  
Db 699 FDDYREEGKSEDTGIIGNIGRVLKARDTIRFELESMPGSDGTWMLDYLWLNRIISGVY 758  
Qy 289 --EPAPTTTKSAPTTKPEPAPTTTKEPAPTTTKEPAP-----TTTKEPAPTTKSAPTTPK 342  
Db 759 VIQAPAIKAAS-----DEPIFTTTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 813  
Qy 343 EPAPTTKPKAPPTTKEPAP-----TPK-EPTPT-TPKEPAPT-TKEPAPTTTKE 390  
Db 814 EPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDE 873  
Qy 391 PAPTAPKKA--PPTTKEPAPT-TPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 447  
Db 874 PTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 930  
Qy 448 PTTTKSAPTTKPEPSPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 493  
Db 931 PSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 990  
Qy 494 APT-TPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 552  
Db 991 TPSETPEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1030  
Qy 553 ----TPEELAPTTPEEPTPT-TPEEAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 606  
Db 1031 DEPTPSD-EPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 1089  
Qy 607 PT-TPKETAPTTPKGTAPTTLKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 664  
Db 1090 PSETPEEPIPTDTPSDEPTPSDEPTPSDEPTPSD-DEPTPSD-EPTPSDEPTPSDEPTPSDEPT 1146  
Qy 665 GTAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 723  
Db 1147 SDEPTPSDEPTPS--DEPTPS-----DEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1195  
Qy 724 SDKPAPT---TPKETAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 779  
Db 1196 SDEPTPSDEPTPSD-EPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 1249  
Qy 780 KSPDESTPELSAEPPTKALENSPKPEGVPTTKTTPAATKPEMITTAKDITTRDURTTPEP 839  
Db 1250 -PSDEPTP--SDEPTP-----SETPEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDE 1302  
Qy 840 TTAAPKMTKETAATTTKTESKITATTTQVTSITTTQDTTPFKITLTKTTLAPKVTITTK 899  
Db 1303 PTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1347  
Qy 900 TITTEIMNKPEETAKPKDRATNSKATTPKQ-KPTKAPKPTSTKPKTPMVRVKPKRT 958  
Db 1348 -----PSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1384  
Qy 959 PTPRKMTS-----TPELNPTSRIBAMLOTTTRPNQOTNSK 995  
Db 1385 PTPPTPTPTSGSGSGSGSGGGGGGGTPTPTPTPS-----KPTSTAPT 1433  
Qy 996 LVEVNPKSDEAGGAGETPHMLLRPH---VFMBE----- 1026  
Db 1434 EIE-EPTPSDVPGAIGGEHAYLRGPDGSRFRBNITRAEAAVIFAKLGADESYGQS 1492  
Qy 1027 -----VTFD-----MDYLPVFNQ 1041

Db 1493 ASPYSLADTHAAWAIAKFAATSGQLFKYDPDGTFFKPDONITRABEATVVLHFLTKVKQE 1552

QY 1042 I-----IINP 1046

Db 1553 IMSKLATIDISNP 1565

RESULT 3

T16251

hypothetical protein F35A5.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004

C:Accession: T16251

R:Leimbach, D.

A:Description: The sequence of C. elegans cosmid F35A5.

A:Reference number: Z18485

A:Accession: T16251

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1274 <LEI>

A:Cross-references: UNIPROT:Q20007; EMBL:U46675; NID:g1166613; PID:g1166621; PIDN:AAB526

A:Experimental source: strain Bristol N2; clone F35A5

C:Genetics:

A:Gene: CESP:F35A5.1

A:Map position: X

A:Introns: 1272/2

Query Match 18.0%; Score 1002.8; DB 2; Length 1274;

Best Local Similarity 23.1%; Pred. No. 3.2e-13;

Matches 330; Conservative 99; Mismatches 309; Indels 692; Gaps 75;

QY 117 TKKKTP-----KPP--VDEAGSLDNGDKVTTPTDTSTQHNVKSTSPKITTAKP 166

Db 2 SRAPPTPIKNPAKKWPPWESVD----- 25

QY 167 INRPSPNPSDTSKETSLSLVNKKETVETKTTNNKQSTDCGKETSASKEQTSIEKTS 226

Db 26 -----EEMEVDDETAPSK-----LEKKPSLR----- 49

QY 227 AKDLAPTSKVLAKPTPKA-----ETTTKGPAULTP-----KEPTP- 261

Db 50 -KD-APTQV--PSPGAPSVPIKNPVKKWKPWEDDEPWEAPAPVPAKKVRDPSK 104

QY 262 TTPKEPASTTPKE-----PTPT-----TTKSAPT 287

Db 105 KVPAPDAPGKKIIMAAKBPETLPVPPPTPVKNPVKKFKAPWDDDEVVDVKDAPTVP 164

QY 288 -----KEPA-----PTTKSAPTTPKEP--APTTPKEP-----PTT 317

Db 165 AKKTPVLKKKEPAAAKPRDPSPKAAAPSKBHDPIVPPPTPIKNPAKKWKPWEDDEVPT 224

QY 318 P-KEPAPTTPKEP-----PTTKSAPT----- 339

Db 225 EIKPEPATRKVPALKKKKEPSTSVKPSVSDPSFTKVPVKKEPVPPTPIKNPTKKWKPW 284

QY 340 ---TP-----KEPAP-----TTPKKPAPTTPKEPAPTTPK 366

Db 285 EDETPEVEVKPPVPEKAPVLKKDPAAPAAKARDPSKAAPKKVFPSSVVP--PTPVK 343

QY 367 EPT-----PTTPKEPAPTK-----EPAPTTP-KEPAP----- 393

Db 344 NPVKYKPPWEDDEVPAEVEKKSAPAKETPVLRKEPEPSTTTPSSDPSPKAAPAVKP 403

QY 394 --TAPKKPAP-----TTPKEPAPTTP-- 412

Db 404 RDSPPKATPLQADPKAQEVPTPVKNPVKKYKPPWEDDEVPEVEVKQPEAPAKKTPLV 463

QY 413 --KEPA-----PTTKP-----SPTTP-KEP----- 432

Db 464 KKEPAAKDTAKFATSKTPTPEKKDPVKPRDPSPKVAAKPDASAQAPATPVKNPVKKWR 523

QY 433 -----PTTKSAPT--TTKEPAP-----TTTKSAPTTPKEPSPTTTKE 468

Db 524 PPWEDDETPADDVSKPTDAKKTPTSLAKKDPAPAKESLKPADTK-APAKPRDPSP--XK 579

QY 469 PAPTTPKEPAPTTPKK---PAPTTPKEPAPTTPKEP-----APTTPKK 508

Db 580 VAPTAPEKKTPLVLAKEPAGPADSKTKEPEKSKPRDPSKKAAPVPAKVPKTEVAPAAVKK 639

QY 509 PAP-----TAPKEPAPTTPKETAPTTP-----KKLTP-----TTEKLAP 543

Db 640 PEPISKPKDTAPKKAEPNSP--VVPPTPVKNPVKKWKPWEDDDAPAKFVSLPEPEKKTTP 697

QY 544 TTPKEPAPTTP-----ELAPTPEPTP--TTPPEEAPTTP----- 577

Db 698 VLAKK-APTKEPDEAAADPVSGSSKDKLAKKAPVPRDPSMKAVPIKPAKTEVPPA 756

QY 578 -----PK--AAPNTPEKPAAPTTPKEP----- 597

Db 757 VVKKEPVAKSRDPSFKKAKAEPNSFVVP--PTPVKNPVKKWKPWEDDDAPAEVNVVPP 815

QY 598 -----APTTPKEPAPTTPKETAPTTPKGTAPTTLKEBAPTTP-PKKAPAPKE----- 641

Db 816 EKKTPVLAKTTPVKPRDPSKKAAPKSTKTDAPPVSVKKEPVPKPEPSPKKAEPNS 875

QY 642 --LAPTTPKEPT-----STTSKAPAPTTPKGTAPT--TPKEP-----APTTPKEP 682

Db 876 PVVPPTPVKNPVKKWKPWEDDDDEPTEEVKPSPEKKTPLVLAKEPEPEKPKDAPKVAAP 935

QY 683 APTTPKGTAPTTLKEBAPTTP-----TPKKPAP-----RELAPTTPTKGPTS-----TTS 724

Db 936 RDESPKKAEP--KEPAKVAAPRDLSPKKAIPANTQOEAAPTTPVKNPVKKWKPWEDD 993

QY 725 DKP-----APTTPKEPAP-----TTPKKPAPTTPPETTPPTTSEVSTP 769

Db 994 DEPAEPVSAPEPEKKTPLVLAKEPAPKPRDPSKKAAPVAAKPDPKIPEVP-----P 1044

QY 770 TTTKEPTTHKSP-----DESTPELSA-EP---TPKALENSPEKEPV-----P 808

Db 1045 TPVKNPVKKWKPWEDDDDEPSEVSAPEPEKKTPLVLAKEAPATKPAKPDSEAAADPVSGP 1104

QY 809 TTKTPAAT-----KPEMTTAKDKTTERDLRTTPTT---TAAPKMTKETATTTKTTES 860

Db 1105 TSKDPLSKKAPVEKPKPTDPKDDKLKSPAKKEKAPAPAAKPKWKPWDDDDPDEPEA 1164

QY 861 KITATTTQVSTTTQDTPPKITTLKTTLTTLAPKVTTTTKTTITTEINKEEATAKPKDRA 920

Db 1165 DFTVPAPSKKPDPTEDPADPLG-----GPKTKDPK-----LNKKAPAEKPTKEK- 1206

QY 921 TNSKATTPKQKPTKAPKKTSTKPKTTPRVKPK-----TTPTPKMTSTM 968

Db 1207 -----PKPEKVSKEPPKPTPEPKP--AAPKKWKPWEDDDDEPEADFTWPAPK----- 1253

QY 969 PELNPTSGRIAEAMLQTTTRENQNTNSKLVEVNPKSSED---AGGAEGETP 1014

Db 1254 -----PTEDPADSLGGGPKPKDP 1271

RESULT 4

T31108

cyst germination specific acidic repeat protein precursor - Phytophthora infestans

C:Species: Phytophthora infestans (potato late blight agent)

C>Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004

C:Accession: T31108

R:Goernhardt, B.

A:Reference number: Z20986

A:Accession: T31108

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1489 <GOE>

A:Cross-references: UNIPROT:O96449; EMBL:AF061185; NID:g3851513; PID:g3851514; PIDN:AAO

C:Genetics:

A:Gene: car90

Query Match		17.8%;	Score 993.7;	DB 2;	Length 1489;
Best Local Similarity		26.3%;	Pred. No. 6e-13;		
Matches 400;		Conservative 57;	Mismatches 422;	Indels 643;	Gaps 74;
QY	27	LSCKGRCFESFERGRC-	-----DC-----	DAQCK-----	KV 52
Db	43	IFCSGVGAEPV--GTACPKAGDVATSDCQPYLLSYNGAVCPVDAECALIHDDMGCEF	-----	-----	100
QY	53	DKCCPDYEFCABEHSVSENQESSSSSSSSSSSSTTWIKSKSNKSANRELQKLKVKDN	-----	-----	112
Db	101	PK--TGYTSAVEAETIAAANGESSGWTGH-	-----	-----	137
QY	113	KN--RTKKPKPPPV--	-----	VDEAGSGLDMDP-	140
Db	138	EEIIPARVNDVTDPIGVNCEVATETATQGHATGGKYDTPSTGTQGTGNTIHYG	-----	-----	197
QY	141	-----KV-----	-----	-----	150
Db	198	STTEGVTKGGYPTDAKVIDGETYLDYPTGITEIIEDGTPPGYGTTDGGTTGGYTT	-----	-----	257
QY	151	-----QHNKVTSPKI-----	-----	-----	161
Db	258	VDNHETTEGAGGYDAGTBEYSTTPVGYSTBETEGQHVGGYFSPDEAPTEGTYV	-----	-----	317
QY	162	-----TTAKPIN-----PRPSLP-----	-----	-----	174
Db	318	PREETAAFSEDTYAPREVTPYAPTEKPYDVBEITYVTEESTYAPTKSETNAPTERMWHY	-----	-----	377
QY	175	-----PNSDT-----	-----	SKETSLTVNKETTVE--TKETT--TTNKQT	205
Db	378	AHIEKPCDTEVTVYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET	-----	-----	437
QY	206	STDGKSKTSKETSQSIKTSKADLAPTSKVLAKPAPKAETTTKGPALTTKPKPPTPT--	-----	-----	263
Db	438	TYAPTEETT-----YAPTEKTTVAPT-----	-----	-----	479
QY	264	-----PKE-----	-----	PASTTPKEPTPTTIKSAPTTPKEPAPT	293
Db	480	APTKEITYAPTEETTYASTBETTYAPTEETTYAPTEETTYAPABETTPYEPTET--TYAPTEETTYAPT	-----	-----	538
QY	294	-TTKSAPTTPKEPAPT--TTKEPAPTTPKEPAPTTPKEPAPTTPKSAPT-----	-----	-----	340
Db	539	EEITYAPTEETTYAPTEETTYAPABETTPYEPTETTYAPTEETTYAPTEETTYAPTEET	-----	-----	597
QY	341	PKPAPPTPKKAPPTTPKEPAPTTPKEPAPTTPKEPAPT-----AP-----TTKEP	-----	-----	383
Db	598	TYAPTEETTYAPABETTPYEPTETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP	-----	-----	657
QY	384	APTTPKEP-----APTAPKKPAPTTPKEPAPTTPKEPAPT--TTKEPSPPTTPKEP-----	-----	-----	431
Db	658	ABETTPYEPTETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPABETTPYEPTET	-----	-----	717
QY	432	--APT--TTKSAPT-----TTKEP-----APT--TTKSAPT--TPKEPSPPTTPKEP	-----	-----	469
Db	718	TYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP	-----	-----	777
QY	470	-----APTTPKEPAPTTPKAPPTTPKEPAPT--TPKEPAPTTPKAPPTAPKEP-----	-----	-----	517
Db	778	TGETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET	-----	-----	837
QY	518	-----APTTPKETAPTTPKLTPTTPKEKLAPT--TPKEP-----APTTPKEKLAPT	-----	-----	559
Db	838	TYTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP	-----	-----	897
QY	560	T-----TPKEPTPTTPKEP-----APTTPKAAAPNTPK	-----	-----	587
Db	898	TEETTYASTBETTYAPTEETTYAPABETTPYEPTETTYAPTEETTYAPTEETTYAPTEET	-----	-----	957
QY	588	EPAPT--TPKEPAPTTPKEP-----APTTPKETAPTTPKGTAP-----TTTKEP	-----	-----	629
Db	958	TYAPTEETTYAPABETTPYEPTETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP	-----	-----	1017
QY	630	APTTPKEP-----APKE-----LAPT-----TTKEPTSTTSKPA-----PITPK	-----	-----	664

Db	1018	ABETTPYEPTETTYAPTEETTYAPTEETTYASTBETTYAPTEETTYAPABETTPYEPTET	-----	-----	1077
QY	665	GTAPTPKEPAPTTPKEPAPTTPKGTAPT--TLKEPAPTTPKP-----APKE-----LAP	-----	-----	713
Db	1078	TVAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPABETTPYEPTETTYAPTEETTYAP	-----	-----	1137
QY	714	T-----TTKGPTSTTSKAPAPTTPKETAPT--TPKEPAPTTPKP-----APT	-----	-----	755
Db	1138	TEETMYAPIEETTYGPTBET--YAPTEATTVAPTEETTYAPTEETTYGPTETTYAPTE	-----	-----	1195
QY	756	PETPPPTTSEVSTPT--TTKEPTTIHKSPDESPELSAETPKALENSPKPVGPTTKP	-----	-----	813
Db	1196	EITYAPTEETTYAPTEETTYAPT--BETTPYEPTETTYAPT--BETTPYEPTETTYAP	-----	-----	1249
QY	814	AATKPEMTTAKDKTTERDLRTPETTTAAPKMTKT-----ATT--TEKTIESKIYA	-----	-----	864
Db	1250	-----TEETTYAPTEETTYAP--TEETMYAPIDETTTYGPTETTYAPTEA	-----	-----	1292
QY	865	TTTQVTSIT--TODTT--PPKIT--LKTITLAPKVTTT-----KKT	-----	-----	900
Db	1293	TVAPTEETTYAPTEETTYGPTETTYAPTEETTYAPTEETTYAPMEETTPYEPABESTST	-----	-----	1352
QY	901	ITTEIMNKPEETAKPKDRATNSKATTPKPQ-----KPKAPKPTSTKKPKMPRVKRP--	-----	-----	955
Db	1353	VSTKPCNTEETDEPTDETPSDPTDEPTDPLP-----TDEPST-----PCD	-----	-----	1401
QY	956	-----KTTPTRK-----MTSTMPEL	-----	-----	971
Db	1402	NOGINGIGENVKRYNNAGIYNTTPGRNSQSWHSCCRSCYNDPICHAFSFHOTS-----	-----	-----	1456
QY	972	NPTSRIAEAMLQTTTRPNQTFN 993			
Db	1457	--SDSVCELTTSTSDREDOQN 1476			
RESULT 5					
I38346					
elastic titin - human (fragment)					
C;Species: Homo sapiens (man)					
C;Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 09-Jul-2004					
C;Accession: I38346					
R;Label: S.; Kolmerer, B.					
Science 270, 293-296, 1995					
A;Title: titins: giant proteins in charge of muscle ultrastructure and elasticity.					
A;Reference number: A57430; MUID:96026330; PMID:7569978					
A;Accession: I38346					
A;Status: preliminary;					
A;Molecule type: mRNA					
A;Residues: 1-7962 <RES>					
A;Cross-references: UNIPROT:Q10465; EMBL:X90569; NID:g1017426; PIDN:CAA62189.1; PID:g10:					
C;Genetics:					
A;Gene: GDS:TTN					
A;Cross-references: GDS:127867; OMIM:188840					
A;Map position: 2q31-2q31					
Query Match					
Best Local Similarity					
Matches 412;					
Conservative 116;					
Mismatches 410;					
Indels 2420;					
Gaps 99;					
QY	14	LSVFVIQOVSSQELSC-----KGRCFE-----	SFERGR-----	41	
Db	4497	LEIFDVDTSQSGEYTCIVSNEAGKASCTTHLYIKAPAKFVKRLNDYSIEGKPIILRGTF	-----	4556	
QY	42	-----ECD-----	-----	45	
Db	4557	TGTPPISVTKKNGINVTSPQRGNITTEKSPILEIPSTVEDAGQYNCYIENASGKDSC	-----	4616	
QY	46	DAQCKKYDKCCPDY-----	-----	59	
Db	4617	SAQILILE---PPYFVKQLEPFPKVSVDGSLQCLAGTPEIGVSWYKGTUKLRPTTYK	-----	4673	
QY	60	-----ESFCABEHSVSENQESS-----	-----	76	

Db	4674	MHFRNNVATLVFNQVDINDSGEYICKAENSUGEVSASTFLTVQEQKLPPSFRQLRDVQE	4733
Qy	77	-----SSS-----	79
Db	4734	TVGLVVFVDFCAISGSEBPISVSWYKDGKPLKDPNVTQTSFLDNTATLNIKFTDRSLAGQYS	4793
Qy	80	-----SSSSSSSTI-----	88
Db	4794	CTATNPISASSARLLITTEGKNPPFFDIRLAPVDVAVGESADFECHVTGTQPIKVSNAK	4853
Qy	89	-----WKIKSSKNSA-----	112
Db	4854	DSREIRSGGKYQISYLENSAHLTVLKVDKDGSGQYTCYAVNEVGKSDCTAQLNIKERLIP	4913
Qy	113	-----KNRRTKKKPT-----	122
Db	4914	PSFTKRLSETVEETEGNSFKLEGRVAGSQPITVAMYKNNIEIQTSNCEITFKNNTLVLQ	4973
Qy	123	-----PKDPVVDE-----	130
Db	4974	VRKAGMNDAGLYTCVSNDAAGSALCTSSIVIKEPKPPVDFDHLTPVTVSEGEVQLSCH	5033
Qy	131	-----AG-----SG-----LDNGDF-----KVITPDT	147
Db	5034	VOGSEPIRIOWLKAGREIKPSDRCSFSGTAGVLELRDVAKADSGDVVCKASNVAGSDT	5093
Qy	148	STTOHKNVSTSPKITTAKPINPRSLPNSD-----TSKETSITVNNKETT-----	192
Db	5094	-----TKSKVT-----IKDKPAVAPATKKAADVGRLPFVSEQPSIRVVEKTTATFIA	5140
Qy	193	-----VETKETTTTN-----	202
Db	5141	KVGGDPIPNVKNWTKWKRQLNQGRVFIHQKDEAKLEIRDTYKDSGLRVCVAFNHEGE	5200
Qy	203	-----KQSTDG-----	209
Db	5201	IESNVNLQVDERKKQKIEGLDRLAMLKKTPIILKKGAGEEBEIDIMELLLKNVDKPEYKYA	5260
Qy	210	-----KEKTTSAKETQSTEK-----	224
Db	5261	RMGIDTFRGLLOAFELLKQSBEEHRLBIEIERSEDERDEKEFEELVSVFIQORLSQTEP	5320
Qy	225	-----TSAKD-----LAPTSK-----	235
Db	5321	VTLIKOJENQVLKONDVAFEDIKINYPEIKLSWYKTEKLBPSOKFELSIDGDRHTLR	5380
Qy	236	-----	235
Db	5381	VKNQQLKDOGNRYRLVCGPHIASAKLTVIEPAWBERHLQDVTLBKGQCTCTMTVQFVSPNVKS	5440
Qy	236	-----	235
Db	5441	EWFRNGRILKPOGRHKTEVEHKVHKLTIADVRAEDQGYTCYKVEDLETSABLRIEAPIQ	5500
Qy	236	-----	235
Db	5501	FTKRIQNIUVSEHQSAFCEVSFDDAIVTWYKGPTELITESQKYNFRNDORCHYMTIHNV	5560
Qy	236	-----VLAKPTPKAETTTKGPALTPPKE-----PTPT-----	262
Db	5561	TPDDEGVYSVIARLEPRGEARSTAEIYLTTKIKLELKPPDIPDSRVPIPTMTIRAVPPE	5620
Qy	263	-----TPKE-----PASTTPKE-----PT	276
Db	5621	EIPFVAVPVLPLPTPEEKKPPPKRIEVTKKAVKDAKKVAKPKEMTPREIIVKPPPP	5680
Qy	277	PTTI-----	280
Db	5681	PTTLIPAKAPEIIDVSSKABEVKIMTITRKKEVQKEKAEVYKQAVHKEXRVFISSFEE	5740
Qy	281	-----	280
Db	5741	PYDELEVEPYTEPEQPYVEEDDEYBEIKVEAKKEVHEEEDFEQEQYEEYEREGYDE	5800
Qy	281	-----KSAPTPPK-----	288
Db	5801	GESEWEEAYOEREVIQVQKEVYEEBHERKVPKAVEKKAPPPKVIKKPVIEKIEKTSRR	5860
Qy	289	-----	288
Db	5861	MEBEKVQVTKVPEVSKKIVPKQPSRTFVQBEVIEVKVPAVHTKQWVISEEKMFASFASHTEE	5920
Qy	289	-----EPAPTTTKSAPTTPKBA-----PTTTKE	312
Db	5921	EVSVTVEVQKEIIVTEEKIHAVNSKRVPPP-----KVPPELPEKPAPEVAVPPIPKKVP	5976
Qy	313	PAPTPKPEPAPTTTKBPAPTTTKSAPTTPKPA-----PTTPKKPAP-----	354
Db	5977	PAPKVPPEVP-----KKDPVEEKKPPVPVKPEAPAPPKVPVEPKKVPVEKIPVPAVKKE	6031
Qy	355	-----TTPKE-----PAPTPKPEPTTPKE-----PAPTTKEPAPTTPK	389
Db	6032	APPAKVEVQKGVVTEBEKIIIVTQREESPPPAVPEIIPKKVPEKVPVPRKEEVEVPPPK	6091
Qy	390	EPAPTAPKKPAP-----	401
Db	6092	VPA-----LPKKVPPEKVAVPVPVAKKAPPPRAEVSCKTVVEBKRFAVEBKLSFAVPQVVE	6149
Qy	402	-----	401
Db	6150	TRHEVSABEESYSSEEBEGVSI SVYREEREEREAEVTEYVMEPEEYVVEEKLHIIS	6209
Qy	402	-----TTPKE-----PAPTTKE-----PAPTTKE--P	423
Db	6210	KRVEABPAEVTREQEKIIVLKPKIPAKIEBPPPAKVPPEAPKKIVPEKKVPAPVPEKKBVP	6269
Qy	424	SPITPKPEPAPTTTKSAPTTT-----KBPAP-----TTPKAP-----TT	457
Db	6270	PPKVPPEKPPVPPEKVPVKVIMKEBPLPAKVIKHEKMLQITQEBKVLVAVTKKEAPKARV	6329
Qy	458	PKPSPTTYKE-----PAPTT-----PK-EPAP-----	479
Db	6330	PEBPKRAVPEKVLKLPKREBPPPAKVTFRKRVVKEEKVSIAPKREPOPIKEVTIME	6389
Qy	480	-----	479
Db	6390	EKEXAYTLBEEAVSVOREEYEEYDYKBFEBEYPTBEBYDQVEEVEEYEEYEREBHEE	6449
Qy	480	-----TTPKKPAPTT-----PAPTTTKKPAPTAPKPE-----	517
Db	6450	YITEBKPFPVKBPEBPPVTKPKAPPAKVLKKAAPKVPVPIPKKLKPPPPKVPPEPK	6509
Qy	518	-----APTTPKETA	526
Db	6510	KVFEKIKH:SI TKREKEQVTEPAAKVPMKPRVVAEBKVPVPRKEVAPPVPRVPEVPEKE	6569
Qy	527	P-----TTPKKLTPTTPKCLAPTTPEK-----	549
Db	6570	PEEVAPEEVEVTHVEEVLVEEBEYIHEBEEFITEEBEVDVPIPKV-PEVPRKPVPPEKK	6628
Qy	550	-----APTTPPELAPT-----TPEEPTPTTPEEPAFT-----	576
Db	6629	PVPVPKKKEAPPAKPAKVPPEVKKPEKVPVLIIPKKEKPPPAKVPPEVKKVPPEKVPVVPK	6688
Qy	577	-----TPKAAAPNTPEKPAPTTPKPEPAPTT-----TPKEPAPTTPKETAPTPKGT-APTTLK	627
Db	6689	KVEAPPAKVPPEVPPKVP-EKKVPVPAKPKKVEAPPAKVPPEVPPKCLIPPEEKPTPVPPKVE	6747
Qy	628	EPAPTTPEKK-----	636
Db	6748	APPPKVPKKREPPVPVVALPQBEVLFEIEIVPEEVLPEEVEVLPEEVEVLPEEVEVL	6807
Qy	637	-----PAP-----	639
Db	6808	EEBIEPPPEEVEVPEEVEVPEEVEVPEEVEVPEEVEVPEEVEVPEEVEVPEEVEVPEEVEV	6867



```

QY 640 -KELAP-----TTTKRP----- 650
Db 6868 KKEEAPPAKVPVKKVEEKRIILPKBEEVLPVEVEEPEEPEEPISEEEIPEBPPPSIEEVE 6927
QY 651 -----TSTTSKAPATPKGTAPT-----PKE--PAPTTPKEPAP----- 684
Db 6928 EVAPPRVDEVKAVDEATPVPKVEAPPAKVSKKIPEEKVDPVQKKEAPPAKVPV 6987
QY 685 ----- 684
Db 6988 KKVPKLVLPKKEAVPAKGRVTLEKVSVAQRQVWVKERLELEVAEVEEIPPEEE 7047
QY 685 -----TPPKG 689
Db 7048 FHEVEEYFEEGEPHEVEEFIKLEQHRVEBEHRVEKVRHVIEWFEABEVEVEFEKPKAPPKG 7107
QY 690 -----TAPTTTL--KEP--APTTPKK-----PA-----PK 709
Db 7108 PEISEKLIIPKPKPTKVVRKPEPPAKVPEVPKIIVVEEKVVRVEEPRVPTKVPVLPK 7167
QY 710 ELAPTTTKGPTSTSDKAPTTPKETAPTTPKEPAPTTPKKAPATTPE-----TPPPTTSE 765
Db 7168 EVVP-----EKKVVPPEAK-----KPEAPPKVPPEAPKEVVPVEKVPVPEPKKPE 7212
QY 766 VSTPTTKE-----PTTIHKSPOESTELSAEP--TPKALENSKEPFGVPTT-- 810
Db 7213 V-PPTKVPEVPKAAVPEKKVPEAIAPPKPEPPPEVEEPEEPPSA---PPKKDEVPVVRV 7268
QY 811 -----KTPAA--TKPEMTTIAKDKTTERDLRTTPTTTTAAAPKMTKETATTTEKT 857
Db 7269 PEVPKEVVPKVPAPAPPKPEV-----TPVKVPEAPK-----EVV 7304
QY 858 TESKLTATTTQVSTTTQDTPPTPKITLTKTLTLAPKVTITTKITTT-----EIM 907
Db 7305 PEKKV-----PVPPKPEVPPTKVPEV-----PKVAVPEKKVPEAIAPPKPEPPPEVF 7353
QY 908 NKPEETA----- 914
Db 7354 EEPEEVALEBPPEAEVSEPEPAAPPQVTPPKNPVPEKPAVAVAKPELPPVKVPVPEVK 7413
QY 915 -----KPKDRAATNSKATTPK-----POKPTKAPKKPTSTK 944
Db 7414 EVVPEKKVPLVVPKPEAPPAKVPVPEVPEKVAVPKPEVPAKVPVPEVKKPVLEE 7473
QY 945 KP----- 946
Db 7474 KPAVVPPEAESEPPPEVVEEPEEIAPEEIEAPKEEKVPVVAESEEPEVPPPAVPEEPKI 7533
QY 947 -----KTMPRVRKPKTTP----- 959
Db 7534 IPEKKVPVIKPEAPPPEPEPEKVBKPKLPRPPPPPPAPPKEDVKEKIFOLKAIPKK 7593
QY 960 -----TPRKMTS----- 966
Db 7594 KVPENQVPEKVELTPLKVPGEKKVKRLIPERKPEKEEVLKSVLRKPEEPEEKVPE 7653
QY 967 -----TWPELNPTSRJAEAMLQTTTR-----PNOTPNSK-----LVEVNPKE 1004
Db 7654 KKLEKVKKPAVPEPPPKPEVEEVVPTVKRERKIPEPTKVPEIKPAIPLPAPPEPKPE 7713
QY 1005 DAGGAEGETHMLLRPHVFMPEVTP-----DMDYLPVFNQGI 1042
Db 7714 -----AEVKT-----IKPPPEVPEPTPIAAPVTVPVVGKKAEPKBAEPKGIKG 7763

```

## RESULT 6

T45025

mucin MUC5B, tracheobronchial [imported] - human (fragment)

C;Species: Homo sapiens (man)

C;Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jul-2000

C;Accession: T45025

R;Desseyn, J.L.; Guyonnet-Duperat, V.; Porchet, N.; Aubert, J.P.; Laine, A.

J. Biol. Chem. 272, 3168-3178, 1997

A;Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alternat

A;Reference number: Z22899; MUID:97166151; PMID:9013550

A;Accession: T45025

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-3570 &lt;DES&gt;

A;Cross-references: EMBL:Z72496; NID:g1834502; PIDN:CAA96577.1; PID:g1834503

A;Experimental source: placenta

C;Genetics:

A;Gene: MUC5B

Query Match 17.5%; Score 975.4; DB 2; Length 3570;

Best Local Similarity 13.2%; Pred. No. 4.6e-12;

Matches 433; Conservative 129; Mismatches 430; Indels 2286; Gaps 96;

QY 6 LPTYLALLLSVFVIOQVSSQELSK-----GRCFESFER----- 39

Db 3 LPV-----STVCREV-----CRWSWNGHRPPEGLGGGDFETFENLRQRGYQVCPV 50

QY 40 -----GRECDGD-----AQCKYDKCCPDYE--SFCAE----- 65

Db 51 LADIECRAAQLDPMLEELGQQVDCDRMGLMCANSQQPPLCHDYELRVLCCEYVPCGP 110

QY 66 -----BHSVSENOE-----SSSSSSSSSSSTIWKIKSKNSAA----- 99

Db 111 SPAGTSPQPSLSASTEPAPVPTQTATKTLWVTPSIRSTAALTSTQSSSGPVTVT 170

QY 100 ----- 99

Db 171 PSAPGTTTCQPRQWTEFDEYPKSEQLGGDVESYDKIRAGGHLCQPKDIECOAESF 230

QY 100 -----NRELOKKLVKDNKQRT-----KKKTPKEPVV 128

Db 231 PNWTLAQGVQKCHDVHGLVCRNWEQGVFKCYNRVIRVLCSDHDCGRATTPPTT 290

QY 129 DEAGSLDNGDFKVTPTDSTTC-----HNKVTSPKITTAKPINP----- 169

Db 291 E-----LETATTTTQALSTPQPTSPGLTRAPPASTAVPTLSGLTSPRY 338

QY 170 -----RPSLPN----- 176

Db 339 TSTLTATTGGPRQSGASTEPTVGVATSTLPTRSALPGTTGSLGTWRPSQPTLAPT 398

QY 177 -----SDTSKE--TSL--TVNKETTVEKETTNTKQTSDDGKEKTSKET- 219

Db 399 ATSRARPTGPTASTASKEPLTSLAPLTSELSTQAEISTPTRETTTNTTTSOGTT 458

QY 220 -----OSIE----- 223

Db 459 RCQPKCEWTEWFDVDPPTSGVASGDMETPENIRAAAGKMCWAPKSIICRAENYEVESIDQ 518

QY 224 -----KTSKADLAP- 232

Db 519 VGQVLTCSETGLTCKNDOTGRFNMCFNVRVLCDDYSHCPSTLATSSTAPSTPG 578

QY 233 TSKVLAKP-----TPKAETTTKGALTPKEPTP----- 261

Db 579 TFWILTPTTTATTASTGSTATASSTQATAGTHVSTTATTPTVTSSK-ATFPSSPGTA 637

QY 262 -----TTPKEPASTTPKEPTPTTIK----- 281

Db 638 TALPALRSTATTATSTFTALPSSSLGTWTRLSQITTPMATWSTATPSTPTVHTSV 697

QY 282 -----SAPT 286

Db 698 LTTTATTGATGVATPSSTPGTAHTTKVLTITTTTGTATPSSSPGRARLPLVMISTTTT 757

QY 287 P-----KEPAPTTT----- 295

Db 758 PTTTRGSTVTPSSIPGTHHTPTVLTTTATVATGSMATPSSSTQTSGGPPSLTTATTITA 817

QY 296 -----KSAP-ITPKEPAPTTT-KEPAPT-----TPKEPAPTTTKEPAPTTKS----- 336

Db 818 TGSTNPSPTGTPPIPVLTATTATPAATSSSTVPSALGTHTPVPVNTTATTHGRSL 877  
QY 337 APTTP-----KEPAPTTPKKPAPT-----PKPAPTTPKEPTPT 372  
Db 878 SPSPHVCTAMTSATSGILGTHITPEFGTSTPAATGTTQHSFALSSPHSRRT 937  
QY 373 PKPAPTTPKEPAPT--TPKBPAPTAPKKPAPT--PKP--APT-----411  
Db 938 ESPSPGTTTGHGTTATSRITATATPKTRTSLLPQOPTSAPITTVTMGCEPQCAWSE 997  
QY 412-----411  
Db 998 WLDYSYMPGSGGDFDYSNIRAAGAVCEQPLGLECRAQAQGVPLRELQGVVECSLD 1057  
QY 412-----PKPAPTTPKBS-----424  
Db 1058 FGLVCRNREQVKMCFNVEIRVFCNCGHCPSTPATSSSTATSPSTPGTWTWILTELT 1117  
QY 425-----PTTPKEPAP-----TTTKSAPT-----441  
Db 1118 ATTTESTGSTATFTSLRTAPPPKVLTTATTPTVTSKATPSSPGTATAPALRSTAT 1177  
QY 442-----441  
Db 1178 TPTATSVTPIPSSSLGTTWTRLSQTTTPTATMTATSPSTPETATHTSTVLTATATTGAT 1237  
QY 442-----TTKEPAPT-----KSAPTTPKEP-----SPT-----465  
Db 1238 GSVATPSSPGTAHTTKVPTTTTGFTATPSSPGTALTTPVWISTTTTTTTRGSTVTPS 1297  
QY 466-----TKEPAPT 473  
Db 1298 SIPGTHTAIVLTITTTTATGMAWPSSSTQTSQTPPSLTITTTATITATGSTINPSSTP 1357  
QY 474 PKEPAP-----TPKBPAPT-----TPKEPAPTTPKEPAPT-----506  
Db 1358 GTRPIPVLTATTATPAATSSSTVTPSSALGTHTPVPVNTTATTHGRSLSPSPHTVTA 1417  
QY 507-----KKPAPT-----APKEPAPTTP 522  
Db 1418 WTSATSGLTGTHITPEFGTSHTPAATGTTQHSFALSSPHSRRTTESPSGTTTT 1477  
QY 523-----KETAPTTPK-----LTPTTPEKLAPTTP-----546  
Db 1478 GHTTATSRITATATPKTRTSLLPSSPTS-APITTVTMGCEPQCAWSEWLDYSYMPG 1536  
QY 547-----EKP-----549  
Db 1537 PSGGDFDYSNIRAAGAVCEQPLGLECRAQAQGVPLRELQGVVECSLDGFLVCRNREQ 1596  
QY 550-----APTTPBELAPTPEE-----564  
Db 1597 VGKFKMCFNVEIRVFCNCGHCPSTPATSSSTATSPSTPGTWTWILTEQTTAAITATGST 1656  
QY 565-----PTPTTPEEPAP-----TTP-----KAAAPNTPK-----EPA 590  
Db 1657 AIPSPGTTAPPPKVLTSQATTPATSSKATSSSPRTATTLPLVLTATKSTATSFPI 1716  
QY 591 PTT-----PKP-----APTTPKEPAPTTPKETA 614  
Db 1717 PSSLTGTTGTSQNRPHPMATWSTIHPSSTPETHTSTVLTATKATTTTRATSSMSTPSSTP 1776  
QY 615 PTT-----PKGTAPT-----LKEPAPT-----632  
Db 1777 GITWILTELTAAATLPHGTPSPSTPGTWTWILTEPSTATVTVTGSTATASSTRATA 1836  
QY 633-----TPKK-----PAPKELAPT-----645  
Db 1837 GTLKVLSTATPTVVISSRATPSSPGTATAPALRSTATTTATSWAIPSSSLGTAWT 1896  
QY 646-----TPKEPTS 652

Db 1897 RLSOTTTPTATMTSTATPSSPTETVHTSTVLTITTTATTTGSAVATPSSPTPGTAHTTKVPTT 1956  
QY 653 TT-----SDKEPAPT-----662  
Db 1957 TTTGFTATPSSSPGTALTTPVWISTTTTTPTTRGSTVTPSSIPGTHATVLTITTTTATVAT 2016  
QY 663-----PKGTAPTTPKEPAPT-----678  
Db 2017 GSMATPSSSTQTSQTPPSLTITTTATGATSTNPSSTPGTTPPPVLTITTTATTPAATSS 2076  
QY 679-----PKP--APT-----685  
Db 2077 TVTPSSALGTHTHPPVNTTATTHGRSLPSPSPHTVPTAMTSATSGILGTHITPESTGT 2136  
QY 686 --TPKGTAPT-----LKEPAP-----TPPKPAPKELAPTTPKPTSTTSDKPAPTTPKE 734  
Db 2137 SHTPAATGTTQSTPALSSPHSRRTTESPPSP-----GTTTPGHTGTSRTTATATATPESK 2192  
QY 735 T-----AP-----TTPKEP-----743  
Db 2193 TRTSTLLPSPSTAPITTVTTTGCPEQCAWSEWLDYSYMPGSGGDFDYSNIRAAGGA 2252  
QY 744-----743  
Db 2253 VCBQPLGLECRAQAQGVPLRELQGVVECSLDGFLVCRNREQVKMCFNVEIRVFCFN 2312  
QY 744 --APTTPKPAPTTPETPPPT-----763  
Db 2313 YGHCPTPATSTATPSSPTGTTWILTKLTATTATTESTGSTATSPSTOGPPAGTPHVT 2372  
QY 764-----SEVSTPTTT-----772  
Db 2373 TATPTVTSKATPSPSPGTATAPALRSTATIPTATSTFAIPSSSLGTTWTRLSQTTTP 2432  
QY 773 -----KEPTTIHKS-----781  
Db 2433 MATMSTATPSSPTETVHTSTVLTITTTATGATGVATPSSPTPGTAHTTKVPTTTTGTFTV 2492  
QY 782-----PD 783  
Db 2493 TPSSSPGTARTPPVWISTTTTTPTTSGSTVTPSSIPGTHTPVLTITTTQPVATGSMATPS 2552  
QY 784 EST-----PELSAEPKALEN-----800  
Db 2553 SSTQTSQTPPSLTITTTATGATSTNPSSTPGTTPPELTITTTATPAATSSVTPSSA 2612  
QY 801-----SPKEP-----805  
Db 2613 LGTHTTPVPVNTTATTHGRSLSPSPHTVTRTAMTSATSGTLGTHITPESTGTSHTPAAT 2672  
QY 806 -GVPTTTKTPAATKPEMT-----TTAKDKTTE-----830  
Db 2673 TGTTTTSTFALSSPHSRRTTESPPSPGTTTTPGHTTATSRITATATATPKTRTSTLLPSQP 2732  
QY 831-----830  
Db 2733 TSAPIITTVTTGCEPQCAWSEWLDYSYMPGSGGDFDYSNIRAAGAVCEQPLGLECR 2792  
QY 831-----830  
Db 2793 ATAQGVPLGELQGVVECSLDGFLVCRNREQVKMCFNVEIRVFCNCGHCPSTPATSS 2852  
QY 831 -----RDLETTTPTT-----TAAPKMTKETA-----TTTEKTTESK 861  
Db 2853 STAMPSSPTGTTWILTELTITTTATGATSTATSPSTPGTAPPPKVLTSPTATPTATSSK 2912  
QY 862 I-----TATTTQV--TSTTTQDT-----TPPKITTKTTLAPKVTTTK-KTITTTIMN 908  
Db 2913 ANSSSPRTATTLPLVLTATKSTATSVTPIPSILGTTGLPEQTTTPVATMSIHPS 2972  
QY 909 KPEE-----TAKPKDRATNSKAT-----TPKPK-----932  
Db 2973 TPETTHTSTVLTATKATTRATSTSTPSSPTGTTWILTELTAAATTTAGTATPSPSTPG 3032



```

Db      1733 TLPSLKEASVLSPTATSSGK-----DSHISPVSDACSGNGTTT--PQASEKLPSK 1779
QY      940 -----PTSTKPKXTWPR--VRK 954
Db      1780 KGTAFTAMLAAPAPESALAITAPIOKSPGANGSASSPKCPDPSSKKDTGLPSAVALA 1839
QY      955 PKTTPTRKWTSTWPELNPSTRIAEAMLTQTTTRNQ----- 990
Db      1840 PQIVPEK-----DISKAETLVSAPKAGDCCLHSPKPGVQGVATPLAAFTSDK 1889
QY      991 -----TPNSKLVEVNPKSEADAGAGETPHMLLRPHVEMPE 1026
Db      1890 VPPEAVSASVAPKPAPAPASLTLPASPVALPPKQPLLESAPGVLESPSKL-----P 1941
QY      1027 VTPDMYLPV-----PNOGIIN-----PML 1048
Db      1942 VPAEDELPLIPPEAVSGGEFPQILVNNMPAKPAGTAPAPASAKQFVL 1991

RESULT 8
S48478
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - Yeast (Saccharomyces cerevisiae)
N:Alternate names: extracellular glucosylase; mucin-like protein MUC1; protein YIR019c
C:Species: Saccharomyces cerevisiae
C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: S48478; A26877; E26877; S27281; J06123
R:Rowley, K.
submitted to the EMBL Data Library, October 1994
A:Reference number: S48478
A:Accession: S48478
A:Molecule type: DNA
A:Residues: 1-1367 <ROW>
A:Cross-references: UNIPROT:P08640; GB:Z47047; EMBL:Z38061; NID:g603997; PID:g763364; GS
R:Yamashita, I.; Nakamura, M.; Fukui, S.
J. Bacteriol. 169, 2142-2149, 1987
A:Title: Gene fusion is a possible mechanism underlying the evolution of STA1.
A:Reference number: A91831; MUID:87194600; PMID:3106330
A:Accession: A26877
A:Molecule type: DNA
A:Residues: 1-242 <YAM>
A:Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525
A:Accession: B26877
A:Molecule type: DNA
A:Residues: 762-1331 <YA2>
A:Cross-references: EMBL:M16165; NID:g172523; PIDN:AAA35015.1; PID:g172526
R:Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.
FEBS Lett. 239, 179-184, 1988
A:Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacchar
A:Reference number: S27281; MUID:89031230; PMID:3141213
A:Accession: S27281
A:Molecule type: DNA
A:Residues: 1-31 <PAR>
A:Cross-references: EMBL:X13857; NID:g4551; PIDN:CAA32069.1; PID:g4552
R:Lambrechts, M.G.; Bauer, F.P.; Marxm, J.; Pretorius, I.S.
Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
A:Title: Muct, a mucin-like protein that is regulated by Mss10, is critical for pseudohy
A:Reference number: J06123; MUID:96323237; PMID:8710886
A:Accession: J06123
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1367 <LAW>
A:Cross-references: GB:U30626; NID:g1304386; PIDN:AAC49609.1; PID:g1304387
C:Genetics:
A:Gene: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458
A:Cross-references: MIFS:YIR019C; SGD:S0001458
A:Map position: 9R
C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase
C:Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein
F:5-21/Domain: transmembrane #status predicted <TM1>
F:1350-1366/Domain: transmembrane #status predicted <TM2>

```

Query Match

16.5%; Score 918; DB 1; Length 1367;

```

Best Local Similarity 25.1%; Pred. No. 1.7e-11;
Matches 337; Conservative 108; Mismatches 437; Indels 460; Gaps 64;

QY      9 YLL-----LLLSVFIQOVSSQELSKGRC-----FE----- 35
Db      118 YLIDNPDTFTATEVYATQVNS-----CQWMENFOIQFEYLOGSAAQYASSWQMG 169
QY      36 --SFEKREGC-----DCAQCKYDKCCPDYDYESFCAEHSVSENG 73
Db      170 TTFDLJSTGCNNYDNOQHSQTDPPGFYWNIDCNC-----GGTKSSTTTSTGS 218
QY      74 ESSSSSSSSSSSTIWKIKSKNSAANRELQKLKVKDNKKNRTKKKPTPKPPVDEAGS 133
Db      219 ESSSTTSSESSTTS----- 235
QY      134 GLDNGDFKVTTPDTSTTOHKNKSTSPKITTAKINP-----RPLSPNNSDTSKETS 185
Db      236 -----STSESTTSSTSSSTSSSTTAPATPTTCTKEKPTPTTCTKEKPT 286
QY      186 TVNKETTVETKETTINKQTSQKKTTSKAKTSQIEKTSKADLAPTSKVLAKPTKAE 245
Db      287 PPHHDTPCTCKKTTTCKTCT-----KKTTPVTPPS--SSTESSAPV-----PTPSS 335
QY      246 TT--TKGPAITPKE-----PPTTPKEPASTTPKEPTPTTIKSAP-----TTPKEPAPT 293
Db      336 TTSSAPVTSSTSSAPVPTPSSSTSSAPVTSSTSSAPVTSSTSSAPVTSSTSSAPV 395
QY      294 TTKSAPTTPKEPAP-----TTTKEPAP-----TTPKEPAP-----TTPKEP 329
Db      396 PTPSSSTSSAPVTSSTSSAPVTSSTSSAPVTSSTSSAPVTSSTSSAPVTSSTSS 455
QY      330 APTTKSAPTTPKEPAPTTPKAPPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 382
Db      456 APVTPSSSTSSAPV-----SSTESSAPV-----PTPSSSTSSAPVTSSTSS 507
QY      383 -----PAP---TTPKEPAPTAPKAPPTTPKEPAPT-----PKEPAPT-----T 419
Db      508 SSAPVTPSSSTSSAP--APTSSSTSSAPVTSSTSSAPVTSSTSSAPVTPSSSTSS 566
QY      420 TKPSPTTPKEPAPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAP 471
Db      567 TPTVTSSTSSAPVTPSSSTSSAPVTPSSSTSSAPVTPSSSTSSAPVTPSSSTSS 626
QY      472 ---TTPKEPAPT--TPKKPAPTTPKEPAPTTPKAPPTTPKAPPTTPKAPPTTPKAP 527
Db      627 TSTTSSAPVTPSSSTSSAPVPT-----PSSSTSSAPVTPSSSTSSAPVTPSSST 682
QY      528 TTPKLTPTTPEKLAPTTPEKLAPTTPEKLAPTTPEKLAPTTPEKLAPTTPEKLA 583
Db      683 VT-----SSTESSAPV-----SSTESSAPV-----PTPSSSTSSAPVTPSS 730
QY      584 NTPKEPAP-----TTPKEPAPTTPKEPAPTTPKEPAPT--TPKGTAPTTLKBPAPT 637
Db      731 ESSAPVTPSSSTSSAPV-----SSTESSAPVTPSSSTSSAPVTPSSSTSSAP 786
QY      638 APKELAPTTTKEPSTSTSK-----PAPTTPKGTAPTTPKEPAPT-----TPKEP 682
Db      787 TESSAPVPT--PSSSTSSAPVPT--PSSSTSSAPVPT--PSSSTSSAPVPT--PSS 843
QY      683 APTTPKGTAP-----TTLKEPAPTTPKAPKLAAPTTLTKGP-----TSTSDKPA 733
Db      844 SSTTSSAPVSSSTESSAPV-----PTPSSSNITSSAPSSIPSSSTTSSSTGTT-- 897
QY      734 ETAPTTPKEPAPTTPKAPPTTPET--PPPTTSEVSTPTTKEPTTIHKSPPDESPE 791
Db      898 -VTPSSKYPGSGQTSVSSSTTTTIVPTKTTTSTVPTTSTTTTTCVSTGNSAGET 956
QY      792 EPTPKALENSKPEPGVPTTKTTPAAKPEMTT-----TAKDKTTERDLRTTPTET--TAA 843
Db      957 GCSPKVITTT-----VPTTTTSTVSTTTTITVCSGTNSAGETSGCSPKITTTV 1011
QY      844 PKMTKETATTBTKTESKITATTTQVTSPT--TTQDTTPPK-----ITLTKVT 898

```

Db 1012 PCSTSPSEASSTTPTPTVTTVSVTVITESTSKPGGEITTTTPTTKNIPTTYLT 1071  
QY 889 TLA--PKVTT-----TKKLTITTEI---MNKPEET---AKPK----- 917  
Db 1072 TIAPTFSVTVTNFTPTTTTTCVSTGNSAGETTSGCSPKTVTTTVPCTGTGTYTTEA 1131  
QY 918 -----DRATNS--KATT-----PKPKTKAPKK----- 940  
Db 1132 TTLVTVAVTTTVTTTSSGTGNSAGTKTTGYTTKSVPTTVTVTLAPSAVTPATNAVPTT 1191  
QY 941 -----TSTKPKTM-----PRVRKPKTTPTPRKMTSTMPBLN- 972  
Db 1192 ITTECSAATNAAGETTSVCASAKTIVSSASAGENTAPSATPTVTTAIPTTVITTESSVGT 1251  
QY 973 -----PTSRIBAM-----LQTTTRNQTPNSKLVENPKSEDA 1006  
Db 1252 NSAGETTTGYTTKSIPTTVTTTLIFGSGAKNYETVATATNP-----ISIKTTSQA 1303  
QY 1007 GGAEGETPHMLLRPHVFMPEVT 1028  
Db 1304 TTASASS-----VAPVVTSPSLT 1321  
  
RESULT 9  
S49915  
extensin-like protein - maize  
C:Species: Zea mays (maize)  
C:Date: 05-Mar-1995 #sequence\_revision 12-May-1995 #text\_change 09-Jul-2004  
C:Accession: S49915  
R:Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.  
submitted to the EMBL Data Library, June 1994  
A:Description: Pex genes: pollen-specific genes with extensin-like domains.  
A:Reference number: S49915  
A:Accession: S49915  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1188 <RUB>  
A:Cross-references: UNIPROT:Q041805; EMBL:Z34465; NID:G600117; PID:G60011  
  
Query Match 15.6%; Score 871.4; DB 2; Length 1188;  
Best Local Similarity 19.4%; Pred. No. 1.2e-10;  
Matches 276; Conservative 90; Mismatches 410; Indels 646; Gaps 49;  
  
QY 10 LLLL--LSVFVIOVSQF----- 26  
Db 17 LLLLAACLSACSQVATSEASVIAHRLAMKEAGGAGOLPADFEEDDRVGAANFN 76  
QY 27 -----LSCKGRCFESFERGECDDAQ--CKYDKCCPD----- 58  
Db 77 PRLRAYIALQAWHRAFYDPKGYTANWGEDVCKYNGVICTEALDDPKITVVAGIDLNG 136  
QY 59 -----YESFCA-----BEHVSQENQ----- 73  
Db 137 ADIAGYLPPELGLTLDAFFHINTNRCFGLIPKMSRLSLILHEFDVNNRNVGVFPYVCL 196  
QY 74 -----ESSSSSSS----- 82  
Db 197 EMVSLKYLDRNDFEGELPPALFDKDLDAIFVNTNRFVGPENLGNSTASVIVFANNA 256  
QY 83 -----SS 84  
Db 257 FVGCIPKSGRMVKTLDLILFLNNKLDGCLPLEWGLLVNTTVIDVSGNMLVGTIPEQLSN 316  
QY 85 SSTIWKIKSKN-----SAANRELQKKLVKDNKKN 115  
Db 317 IAKLEQLDVSRRVFTGIVHESICELPALVNFSAFNFNFSEAAVCMPSDKALVNLDDRN 376  
QY 116 -----RTKK-----PTKPPVWDEAGSLGNDGFKVTTPDT 147  
Db 377 CLGALRPAQKALQCAPVLARPVDCSKHVCAGYETP-----GGG-----PPS 418  
QY 148 STTQHNKVSSTPKITTAKPINPRPSLPNSDTSKETSITVKNKETTETKTTTNKQTST 207

419 SPVPCPKAASADM-----PSPHTPP--DVSPPEP----- 444  
QY 208 DGKEKITSAKETQSTIEKTSAKDLATSKVLAKPTTAKAETTTKGPALTTPKEBPTTP-KE 266  
Db 445 -----LPEPGFVPAPAMPMTTPHSPPADDVVPPTPPVPGKS 481  
QY 267 PASTTP-----KEPTPTTIKSAPT-----TPKEPAPTTTKSAPT--TPKEPAPTTTK 311  
Db 482 PPATSEFSQVOPPAASTPPPSLVKLSPPQAPVGSPPPVKITSPPAPIGSPFPFPVSVV 541  
QY 312 EPAPTTPKPAPTTTKKPAITTKSAPTTP--KEPAPTTPKKPAITTPKPAITTPKPEPTTP 370  
Db 542 SPFPVKSPPPPAPVGSPPPEKSPPPAPVASPPPVKSPBPPTLVASPPPVKSPPPP 601  
QY 371 TTPKPAITTKPAPTTTP--KEPAPTPAKKPAITTKKPAITTPKPAITTPKPEPTTP 425  
Db 602 APVASPPPPVKSPPPTTPVASPPPPAPVASSPPPMKSPPPPTPVSSPPPEKSPPPPPA 661  
QY 426 --TTPKPAITTKSAPTITTKPAPTTTKSAPTTPKPEPTTKKPA--PTTPKPAITTP 482  
Db 662 KSTPPPEYPT-----PPTSVKSSPPPEKSLPPPTLIFSPPOBKPTPSTESKP-PSSP 715  
QY 483 KKPAPTTKPA-----PTTPKE-----PART 504  
Db 716 EKESP--PKEPVSPPTPKSSPPPEAPVSSPPPTPVSSPPALAPVSSPPSVKSSPPPAFL 773  
QY 505 TTKKAPATAPKPAPT-----TPKETAPTTPKLTPTTPKLAPTTPKEKAPITPEEL 557  
Db 774 SSPPPAPOVKSSPPPVQVSSPPPAKSSPPLAP-----VSSPPQVEKTSPPAPLSSSPL 828  
QY 558 AP-----TTPPEPTPTTPEAPTTTPEAPTTTPEAPTTTPEAPTTTPEAPTTTPEAP 599  
Db 829 APKSSPPHVSSPPPVKSSPPPAVSPPLTPKPAHPASPAHVSPPEVVKSTP--PAP 886  
QY 600 TTPKPAITTPKETAPTTP-----KGTAPTTLKEPAPTTPKKPAKELAPTTTKEPT 651  
Db 887 TTVISP--PSEPKSSPPPTVSLPPPIVKSSPPPAVSSPPMTPKSSPP-----PVVSSPP 941  
QY 652 STTSUKPAPTTPKGTAPTTKPAITTPKE--PAPTTPKGTAPTTLKEPAPTTPKKAPKE 710  
Db 942 PTVKSSP--PPAPVSSPPATPKSSPPPAVNLPPPEVKSSPPPTPVSSPPPAKSSPP-- 998  
QY 711 LAPTTTGTSTTSOKPAPTTPKETAPTTKPAITTPKPAITTPKPAITTPKPAITTPKPAIT 770  
Db 999 --APMSGPPPEVKS--PPPPAPVSSPPPVKSGPPPPAPVSSPPPVKSSPPPAVSSPP 1055  
QY 771 TTKEPTTIHKSPDESTPELSAETPKALENSKPEGVPTTKTPAATKPEMTTAKDKTTE 830  
Db 1056 PVKSSPP--PAPISGPPPVKSGPPPAVSSPP-----PPVKSGPPPPAP----- 1097  
QY 831 RLRLTPTTITTAAPKMKETATTTEKTTESKITATTQVTSITTTQDTTTFKLTITL 890  
Db 1098 ----- 1097  
QY 891 APKVTITTKTITTTTEIMNKPEETAKPKDRATNSKATTPKPKPTKAPKPTSTKPKTMP 950  
Db 1098 -----VSSPPP--PIKSPPPAPVSSPPAP 1121  
QY 951 RVKPKTTTPPKMTSTMPBLNPTSRIBAMLTQTTTRNQTPNSKLVENPKSEDA-- 1007  
Db 1122 --VKPSLPPAPVSSPPPVVTPAP-----PKKEQSLPP 1154  
QY 1008 GAGETPHMLLRPHVFMPEVTDMDYLRVFNQGIINPMLS 1049  
Db 1155 PAESOPP-----PSFN-----DILPPIMA 1174  
  
RESULT 10  
T25697  
hypothetical protein F16F9.2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T25697



QY 92 ----- 91  
Db 686 AGSGSGLTPTTLEPKIEGSKKASGVWTEEDGDEDLMEGSGSWTTINGTIGTS 745  
QY 92 -----KSSKNSAANRELQKLV-KDNK----- 113  
Db 746 PRSEGTIRVRITTLGDBGEPETATKPGISAPDKTGESKTESDGEKLTVEKDGEAQSS 805  
QY 114 -----KNRKYKKTPPKPPVDEAGSL-----DNGDFK 141  
Db 806 GSATSGKKGSEATSGSSSSAKSGTGSEAGSSGSSGSGVSGESVSTESGSG 865  
QY 142 VTT-----PDTSTTHQNKVST-----SPKIT 162  
Db 866 TSSSGSVGEATSGTGDGSESGKPSKSTEEKLPFTKNGEKSPISGSDTTGKSEBEIT 925  
QY 163 TAKPNRPSLPNS-----DTSKETS LVNK----- 189  
Db 926 SRKPIRGSLSLTREGSGSGEWFETGSKGHFESGSKSVTSGKPTQSGAEGSGGPKVPKG 985  
QY 190 ----ETTVETKETTT-----NKQTSDDGKEKTTSAKETQ 220  
Db 986 PGAPETITDGERSSSTSGDKSGGKADKSDNNKVPKTDGKNPDITTDGEDSTSETSGGE 1045  
QY 221 SIEKTSADLAP-----TSKVLAKPT----- 241  
Db 1045 QGPKGSKGQPPGDGSEVKKPTSEVDGPNLSGKGNVPLKPTDLP EEGSGILITSS 1105  
QY 242 ----PK-----AETTKGPALTPPKBPTP--TPKE----- 266  
Db 1106 GGNKSTFEHGTCLERLPFKTEDKSSBTPQLGLEISAGKPEPDGTSKEVBLEIWESTT 1165  
QY 267 PASTT----- 271  
Db 1166 POSTILDSVGLIESSDLTKATKPHVEIEGSGTGDEIEITATRDVSKSTKKRPREVD 1225  
QY 272 ----PKEPTTTIKSA-----PTT-----PKEP----- 290  
Db 1226 GGDNGETSGVGGKPTTAPTPSSAESSIRIPTTSEAPSGSGEAGVPESPDGSGESS 1285  
QY 291 ----ATTTKSAPTTKEPAPT-----TKEPAPTTPKEPAPT----- 325  
Db 1286 TSAPDGVSTSSATAPEVTTASSTPDVAEESGIPSTKPTAEPLTT--APSTEV 1341  
QY 326 ----TKEPAPTTKSAPTTPKEPAPT-----KKAPAPTTPKEPAPT----- 364  
Db 1342 SPGSGTEESTLPPTEGSGESTTSSAPT--EPATVLPQNRNEKPEPTKDTFALPTTT 1399  
QY 365 ----PKE----- 367  
Db 1400 GAPQANDSSVENTKCTSSDBEGLDALCERRFTGVCRCPEGAPPKSCVDVDECATGDH 1459  
QY 368 ----PT----- 369  
Db 1460 NCHESARCONVGGYACFCPTGFRKADGSCQDIDECHEHNSCCGANAKCVNKPCTYSC 1519  
QY 370 ----PTTPKEPAPTTK----- 381  
Db 1520 ECENGFLBGYQCVPTT-KKPCDSTQSSKSHCSNMSCEVDTVDSGVECKCMGGYKKS 1578  
QY 382 ----EPAP----- 385  
Db 1579 GKVCEDINECVAEKAFCSLNANCVNMGTFSCSKQGYRGDGPMTDIDNECDERHPCPH 1638  
QY 386 ----TPKE----- 390  
Db 1639 AECTNLEGSFKCEHSGFEGDGIKKCTNPLERSCEDVEKFCGRVDHVSCLSVRYNGSL 1698  
QY 391 ----PA----- 392  
Db 1699 SVCEBPGFRFEKESNCVDICEBSRNNDPASAVCVNTEGSKYRCECABEGYGGVC 1758  
QY 393 ----PTAPK----- 397

Db 1759 TDIDCDRGMAGCDSDMAMCINRMGSCGCKMAGYTGDGATCIKIBEEPKSDKTADEWS 1818  
QY 398 ----KPAPTTPKEPAP----- 409  
Db 1819 RLCELEKKQCTVDEERVPOCGACLPGHHPINGTCQSLQISGLCAQKNDCKNHAECIDIHP 1878  
QY 410 ----- 409  
Db 1879 DSHFCSCPDGFIDGMI CDDVDECNAGMCDDEKTCENTIGSFNCVLEGGFKKVDKCV 1938  
QY 410 -TTPKEP-----APTTKEPSPPTTPKEPAPTTPKSAPTTPKBPAPTTKSAPT 456  
Db 1939 VDEKKQPNREKIEIDRENSSSNGGKPTTKGIVSSTSATSSSESTAPHVHTIISST 1998  
QY 457 TPKEPSPPTTPKEPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTAPKE 516  
Db 1999 STKDMTSSKSPENVMTMSSEPEVSTSSSKSTTASSETTVSSTPSESS-----SSEAPLT 2051  
QY 517 PAPTTPKETAPTTPKLTPTTPE-----KLAPTTPE-----KPAPTTPPEELAPT 561  
Db 2052 SSPATTTEV--ITESVSKSTTPKESSSEITVKLSKSPSEVTSSESSVSKSPSTPS-----TT 2105  
QY 562 PBEPTPTTPEEPAPTTPKAAAPNTPKEP-----APTTPKEPAPTTPKBPAPTTPKET-- 613  
Db 2106 SQSVTSTVPETSKSTVLSSEAPVTSPTSEVHTSSETKSLSASSTGDNSTTPSTSSL 2165  
QY 614 ----APTTPKGT-----APTTLKEPAPTTPKPKAPKELAPT-----TTK 648  
Db 2166 ASVKSTSAPGTSASVAPVKLSLSDV--SQPSTKTFDATESTVQASSETSSOTS VKSTS 2224  
QY 649 EP-----TSTTSDKBPAPTTPKGTAPTTPKEPAPTTPKBPAPTTPKG----- 689  
Db 2225 EPESHVKLSITSSNPSVPTSPKSTPTPEST--EQFTSTTPSGQSLTPMNSNEVL 2282  
QY 690 ----TAPTTLKEPAPT--TPKKAPKEL--APTITTKGPT----- 720  
Db 2283 TTSEPHVLSLSDVDSQSTTPNNLSESTVETPKTSSEVLSNSEPSTTEAPTILSPD 2342  
QY 721 -STT-----SDKP-----APTTP----- 732  
Db 2343 ILSTTTNNLSQSTVSTEDRSSESENSEKPTSAPELVTSSVTHVASSPDVPTESSEP 2402  
QY 733 ----KETAPTT-----KEPAPTTPKP-----APTPE-- 757  
Db 2403 DLTGSSTENIPEASSKQITISPTPTDTTASEPTKSTSMSPDLSSTSNVLSSESSTPES 2462  
QY 758 ----TPPPTTSEVSTPTTKBPT 777  
Db 2463 SSKSPVSSSTEGISVVVSTSEFSKVPESTISSVLEEDLTKTTPSPILEETTTASETSEPLT 2522  
QY 778 ----IHK--SPDESTPELSABPTPKALENSP-KEPG-----VPTT----- 810  
Db 2523 EDSLATVSVRIHELTTSSENVKPESESTTTSSSESKPEQPAGILTSTVVVPTSSVSLITA 2582  
QY 811 ----KTPAATKPEM-----TTAKDKTTERDLARTPETTTAAPT 847  
Db 2583 SEIEALTSNTPPKQGRPTITTPSKLSVKSTTSPSTVTSSEPESESTKRTTSTVSTTTPT 2642  
QY 848 KETATT--TEKTTESKITATTITQVTS--TTTQD 876  
Db 2643 EBITTSESILLTAAPSKPTESSTESSEAPTTPAKTSEKPSNVSTSRKSTENVEITSQ 2702  
QY 877 TTPFKITLTKTT-----TLAPKVTTTKKTTITT----- 904  
Db 2703 SGLSESTTMSSTSBETNAPAVTVSSSEASTTLEENSTSSPTSSEASVKLSLPPESI 2762  
QY 905 -----EIMNKPKEE----- 912  
Db 2763 TSEAVTVSSRAPAEITMSESHREISTVSSPESEPEIPLSTTVSPNVVTASSIPSEPI 2822  
QY 913 ----TAKPKD-----RATNSKA-----TTP 928

Db 2823 SSVTSSTPRVRLITCTPDDLIIVSVTPSHGNRRQNTASSVPSNSTSPILLPSESILTP 2882  
Qy 929 KPQKPTKAKPKTSTTKKPTMPVRKPK-----TTPTP-----RKMTSTMP 969  
Db 2883 QPPTTTTAKPATTSGKRPSPQIOPPAEMFTTTPAPPSPNGGYGEETNOEBQVIST-- 2940  
Qy 970 ELNPTSRIFAMIQT-----TTPRQTPNSKLIVVNPKS 1003  
Db 2941 -----TTTEPISLCSTVTCHSLATCEOSTGVCICRDPGIGDGTACSKKSTADCSIPSLC 2996  
Qy 1004 EDAGGAEGET-----PH--MLLRPHVMPPE 1026  
Db 2997 ADKAKCDNSTRCECDAGYIGDGVVCSHPQDCVLRDNLCSPE 3039  
RESULT 12  
T29757  
protein UNC-89 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 03-Dec-1999  
C;Accession: T29757  
R;Du, Z.; Le, T.T.; Wilson, R.  
submitted to the EMBL Data Library, May 1997  
A;Description: The sequence of C. elegans cosmid C09D1.  
A;Reference number: Z20679  
A;Accession: T29757  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-6642 <DUZ>  
A;Cross-references: EMBL:AF003131; PIDN:AA854132.1; GSPDB:GN00019; CESP:unc-89  
A;Experimental source: strain Bristol N2; clone C09D1  
C;Genetics:  
A;Gene: CESP:unc-89  
A;Map position: 1  
A;Intons: 17/2; 108/3; 154/2; 211/2; 265/3; 326/2; 352/3; 426/2; 454/1; 500/1; 537/1; 6  
/3; 5917/1; 6027/1; 6061/3; 6153/2; 6515/1; 6552/3; 6609/1  
Query Match 15.4%; Score 857.5; DB 2; Length 6642;  
Best Local Similarity 16.5%; Pred. No. 2.4e-099; Indels 1295; Gaps 83;  
Matches 366; Conservative 159; Mismatches 399;  
Qy 9 YLLLLLSVFIQVSSQE-----LSCKGRCFESFERGEC---DCDAQCKK 51  
Db 393 YSSIRLDKYNIRQHTTDEDIVLQPOEPGLPSRIKPKDFTSEYVRKAWLRDIAEBQEK 452  
Qy 52 YDKCCPDYEFCAEHSV-----ENQESSSSSSSSSTIIV----- 89  
Db 453 Y-----AAERDAISMTATSEMTASSVDPMNASDQQSFSEWSGSRKSLFPGPE 502  
Qy 90 -----KIKS-----SKNSAANRELQK 105  
Db 503 EGGPRKKVKSPPVISPTGSSTSIYSGGSSSIDWTTTGTLEMQGTRVTRQYGFRLQE 562  
Qy 106 ----- 105  
Db 563 SSAMKMLKVTGYPDPITWYKDDVQLHEDRHTFYSDEGFFAMTDPVQVDTGRYTCM 622  
Qy 106 ----- 105  
Db 623 ATNEYQASTSAFFRVLKVEEAAPFAVTKLRDKECEGDIVDFECEVGEWPEPELWVL 682  
Qy 106 -----KLKVKDNK----- 113  
Db 683 VDDQPLRPSHDFRLQDVGQTAKLEIRDAQPDGTVTVTKIQNEBFGSIESKAELFVQADPD 742  
Qy 114 KN----- 115  
Db 743 KNHVAPFQATIEYVECEDEGEVRFKSVITGDPNPELIWFINGKPLSESKVKFISBDGI 802  
Qy 116 -----RTKKKPTPKPPVVDE----- 130  
Db 803 CIITIKDVTBFDGMVTCQGSNRLGASCDGRUKVRVPPAPPTFNKPLEDKTVQEKSTVV 862

Qy 131 -----AGSGLDNGD-----FKVTTTPTDSTTOHN----- 153  
Db 863 FEVDVSGNPEPTLFTLCKGKELXNGEGVEIVGHDGFYRISIENTSMKXHDGELVAKAQN 922  
Qy 154 -----KVSTSPKITTAKPI-----NPRPSL----- 173  
Db 923 EHGTAESRARLTVEQEEESRSAP--TFLKDIEDQTVKGFPAVFETTVRGNENPEVTWF 980  
Qy 174 ----- 173  
Db 981 INGHKMDQSGPGVKIBAHNHDKLITDSAQYAGTVLCRAENAVGRFETKARLVVLAPEKQ 1040  
Qy 174 --PPNSDTSKETSLSLVNKKTTV-----ETKETT----- 200  
Db 1041 KKPP-----KFVEILLVDKTTETVDNTVVFVRVEGEPKPTVWYLKGEELKQSDRVEIREF 1095  
Qy 201 -----TNKQSTDGKEKTTSAKETOSIEKTSIAKDLAPTQSKVL 237  
Db 1096 DGSIKISIKNIKIEDAGEIRAVATNSEGSDTKAKLTQVKKPAPE-----FDLRPVSLTV 1151  
Qy 238 AK-----PTP-----KAETTTKGPALTT----- 255  
Db 1152 EKGSEAVFSAHAFGIPLPTYWMSVNGRKVRDQOGARVTRDESTVDGASILTIDTATYYS 1211  
Qy 256 -----PKEPTPTTPKEPAST-----TPKE 274  
Db 1212 EVNHLTISVVAENTLGAETGAQLTIEPKKESVVVEKQDLSSEVQKEIAQQVKEASPEA 1271  
Qy 275 PPTTTIKSAPTTTPKEPAPTTT-----KSAPT----- 300  
Db 1272 TTTTITMETSITSTKTMTMSTTEVTSTVGTVVTKESSESESAITVIGGGSGGVTEGSIV 1331  
Qy 301 -----TPKEPAPTTTKEPAP-----TTPKE-----PAPTTTKEPAPTTTK 335  
Db 1332 SKTEVVSQKDSQDQVREGTPKRRVSAEBELPKVEIDSRKKKSSP--DKKESPKPTE 1390  
Qy 336 SAPTTTPKEPAPTTPKKAPTTTPKEPAPTTTPKEPTTPTPKEPAPTTTPKEPAPTT--KEPAP 393  
Db 1391 EKASPTKTKGEEVKSP--KEKSPASPTKKEKSPAAEEVKSPTKKEKSPSSPTKKEKSP 1447  
Qy 394 TAPKAPAPTTTPKEPAPTTTPKEPAPTTTKGSPSTTP-----KEPAP-----T 434  
Db 1448 SSPTKTKTGDEVKESP--PKSP--TKKESKPEKPDVKSPVKKESPDATNIVEVSSET 1502  
Qy 435 TTKSAPT-----TKEPAPTTTKSAPTT-----KEPSPTTTKEPAPTTTP 474  
Db 1503 TIEKTEMTMTMTHSESESRTSVKKEKTPKVEDEKPKTKKDKSPKESITBEIKSPVK 1562  
Qy 475 KEPAP--TTPKAPAPTTTPKEPAPTTTPK-----PAPT-----A 513  
Db 1563 KEKSPKEVEKPPASPTKKEKSPKASP--TKKSENEVKSPTKKESPEKSVVEELKS 1618  
Qy 514 KPAPAPTTTPKAPAPTTPKKLTPTTPKAPAPTTTPKAPAPTTTPKAPAPTTPEPTP 567  
Db 1619 PKEKSP-----EKADDPK--SPTKKEK-----SPEKATGDKSPKKEKSPKVEEKPTS 1668  
Qy 568 TTPKEPAPTT--TPKAAAPNTPKEPAPTTTPKE--PAPTTTPKEPAP-----TTPKAPAPTT 616  
Db 1669 PTKKESPTKUDDEVKSPTKKESQTVVEEKASPTKKEKSPKSVVEVKSPKESKSP-- 1727  
Qy 617 TPKGAPATTLKEPAPTTTPKAPAPKELAPTTT-----KEPTSTTSDKAPAPTTTPK 664  
Db 1728 -----EKAEKPKSPTKKESPEKSAAEVKSPTKKESPEKSAEEKPKSPKTK 1776  
Qy 665 GTAPT-----TPKEPAPTTTPKEPAPTTTPKGTAPTTTLKEPAPTTTPKAPAPKELAPT 714  
Db 1777 ESSFVRMAADDEVKSPTKKESPEKVEE--KPASP-----TKKE--KTPKESAAEELKSP 1826  
Qy 715 TTK-----GPTSTTSD-----KPAPTTTPKAPAPTTTPKAPAPTTTPKAPAPTTPE 757  
Db 1827 TKKEKSPSPTTKTGDESKESPEKPEEKPKSPTPKSPGSPKPKKSKSPKAEKPPAPK 1886  
Qy 758 TTPPTTSEVSTPT-----TTKPTTIHKSPD----- 783



Db 1887 ----LTFDLKQVWNTDLAHPFVWVVEHATECKWFLDGKEITTAOGVTVSKDQDFRCS 1942  
QY 784 -----ESTPBLSEPTPKALENSPKPGVPTTKTPAATPEMT 821  
Db 1943 IDTMFGSGTVSVASNAAGSVETKTLKLETPK-----ETK-----KPEFT 1985  
QY 822 TTAKD-----KTERD----- 832  
Db 1986 DKLRDMVTKGDTVQMDVIALHSPLYKWYQGNLLEDGKNGVYIKNEENKSSLIIPNAQD 2045  
QY 833 -----LRTTPETTTA-----APKMT-----KETATITTEKTES----- 860  
Db 2046 SGKITVEASNEVGSSSSAQLTVNPSSTTPIVVDGPKSVTIKETAEFAFKATISGPPAPT 2105  
QY 861 -----KITATTQVTSITTTQDITPKKIT-----TLKTT----- 888  
Db 2106 VKWTINEKIVEESRTITTIKTEDVYTLKISNAKIEQTGVTKVTAQNSAGDSQADLKVE 2165  
QY 889 --TLAPKVTTTKTIT----- 2165  
Db 2166 PNVKAPKF---KSQLDKVADEGEPLRWNLLEDGSPGTEVSWLLNGQPLTKSDTVQVVD 2222  
QY 909 -----KPEE-----TAKPKDRA-----TNSKAITPKQKPTKAPKPKTS----- 942  
Db 2223 HGDGYHVTLAEAKPEMSGTLTAKAKNAAGECETSAKVTVNGGKPKPEFVQAPQNHETTL 2282  
QY 943 -----TKYP-----KTMPPRVKP----- 955  
Db 2283 EESVKFSAIVTGKPMENVWYLNKKLIQSEEVKVYVHETGKTSIRIQKFLMEHNGTIR 2342  
QY 956 -----KTTPPKMTSTM----- 970  
Db 2343 VEAEVSGKVQATAQLKVDKKEVTP-KFTTNMDRQKGEDEVKFTANVEGYEPPSVAWT 2401  
QY 971 LN--PTSRIAEMLOTTTPRNPQNSK-----LVEVNPKSEDAAG----- 1008  
Db 2402 LNPGEVSK-----HPNITVDKGHTTEISAVTPEQAGELSCAETNPVSGSKRD 2451  
QY 1009 -----AEGETPHMLRPHVFM-----EVTPEMDY 1033  
Db 2452 VQLAVKKVGDAPTFAKNLEDRLITEGELTMDAKLNIKPKKITWLKOGVEITSDGHY 2510

RESULT 13  
B48666  
cell proliferation antigen Ki-67, short form - human  
C:Species: Homo sapiens (man)  
C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 15-Mar-2004  
C:Accession: B48666  
R:Schlueter, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Gerde  
J. Cell Biol. 123, 513-522, 1993  
A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, ubiqui-  
ous.  
A:Reference number: A48666; PMID:94043435; PMID:8227122  
A:Accession: B48666  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-2897 <SCH>  
A:Cross-references: EMBL:X65551  
A:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat  
F;29-91/Domain: Kinase interaction domain homology <KIH>

Query Match 15.0%; Score 835.4; DB 2; Length 2897;  
Best Local Similarity 15.3%; Pred. No. 2.1e-09;  
Matches 392; Conservative 138; Mismatches 420; Indels 1521; Gaps 94;  
QY 22 VSQELSLCK-----GRCFESFERGECDCD 46  
Db 382 ISNQKDFEDLSGIAEMFPTVKEQPOLSTCHIAISNENLLGQFQGTDSGE----- 437  
QY 47 AQCKYDKCQPDYEST-----CAEEH-SVSE----- 71

Db 438 -----PLLPSTSESGNVFFSAQNAAKOPSKDCASPLRRQCIRENGNVAKTPRNTY 490  
QY 72 ----NQSSSSSSSSSSSTIWKIKSSKNSAANBELQK----- 105  
Db 491 KMTSLETKSDTETEPSKIVSTVNRSGRSTEFNRNIQKLPVESKSEETNTEIVECILKRGQ 550  
QY 106 ----KLKVKON-----KNET-KKKPTPKPPVUDEAG 132  
Db 551 KATLLQORREGEMKEIERPETTYKENIELKENDEKMKAMKRSRTWGQCAPMSDHTD--- 607  
QY 133 SGLDNGDFKVTTPDT-----STTOHNVSTSPKITTAKPINRPSLPP---N 176  
Db 608 ----LKSLEPDTLMKDTARGQNLQTDHAKAPKSEKGIK---MFCQSIQPEPIN 657  
QY 177 SDTS-----KETSILTVMKETTIVETKETTNTNNKOTSTGKE----- 211  
Db 658 TPTHTKQQLKASLGKVGVEELLAVGKFTPT-SGETHTHREPADGCKSIRTFKESPKQI 716  
QY 212 ----KTTSAKETQSIKTSKDLAPTCKVLAKPTKAE----- 245  
Db 717 LDPAARVTGKKWPRTPKEAQSLE-----DLAGFKELFOTPGPSESMDEKTTKIACK 771  
QY 246 ----TTTK-----GPALTTPKEP----- 259  
Db 772 SPPPEVDPTPTSKQWPKSLRKADVEEFALRLKLTTPSAGKAMLTTPKAGGDEKIDKAF 831  
QY 260 --TPT-----TPKEPAST-----TPKEP 275  
Db 832 MGTVPQKLDLAGTLPKSRQLOTPKEKAQALEDLAGFKELFOTPGHTEELVAGKTTKIP 891  
QY 276 ----TPTTIKSAPT-----TPKEP----- 292  
Db 892 CUSQSDPDVDTPTSTQPKRSIRKADVEGELLACRNLMPSAGKAMHTPK---PSVGECK 948  
QY 293 ----TTTKSAPTTPKEPAPT----- 308  
Db 949 DIIIFVGTVPQKLDLENLTGSKRRPQTPEKAQALEDLTGFKELFOTPGHTEEAVAAGK 1008  
QY 309 TTKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKE----- 343  
Db 1009 TTKMPCESSPPESADT-----PTSTRQPKTPLEKRDVQKLSALKLTQTSGETHTD 1062  
QY 344 ----PAPTTPKAPPTTPKEP----- 361  
Db 1063 KVPGEEDKINAFRETAKQKLDPAASVTSKGRHPT-KERQPLEDLAGNKLQFQVCT 1121  
QY 362 --PTTPK-----EPTPT-TPKEPAPTKE----- 382  
Db 1122 DKPTTHEKTTKIACRSQDPDVTPTSSKPSQSKSLRKVDVEEFFALRKETPSAGKAMHT 1181  
QY 383 PAPT-----TPKEP----- 392  
Db 1182 PKPAVSGEKNIYAFMGTVPQKLDLENLTGSKRRRLQTPKEKAQALEDLAGFKELFOTRGH 1241  
QY 393 ----PTAPKPPA----- 403  
Db 1242 TBESMTNDKTAQVACKSSQDLDKKNPASSKRRRLKTSLGKVGKVEELLAVGKLTQTSGETT 1301  
QY 404 PKEPAPT-----TPKEP-----P 417  
Db 1302 HTHTEPTGDKGKMAFMESPKQILDASAALTSKGRQLRTPKGSKEVPEDLAGFIELFQTP 1361  
QY 418 TTTKPESPPTPK-----EP-----APTTPKSAP----- 440  
Db 1362 SHTK-SMTNEKTTKVSRAQPDLDVDTPTSSKQPKPRSLRKADTEEBFLAFRKOTPSAG 1420  
QY 441 ----TTTKEP----- 447  
Db 1421 KAMHTPKPAVGEKDNITFLGTPVQKLDQPGNLPQSNRRLOTRKKAQALEELTGPRELF 1480  
QY 448 ----PT-----SAPTTPKEPSPTTKE-----PA 470  
Db 1481 QTPCTDNPTADEKTTIKILCKSPQSDPADTPTNTQPKRSLKADVEEBFLAFRKLITS 1540

QY 471 -----PTTPKEPAPT----- 480  
Db 1541 AGKAMHTPKAAVGEEXDINTFVGTPTVEKLDLLGNLPGSKRRPQTPKEKAKALEDLAGFKE 1600  
QY 481 ---TP-----KKPAPTTPKEPA----- 494  
Db 1601 LFTQPGHTEBMTDDKI TEVSCSKSPQDPVKTFTSSKQRLKISLGKVGUVKEEVL PVGKLT 1660  
QY 495 -----PTTPKEPA----- 502  
Db 1661 QTSCKTQTHRETAGDGKSIKAFKESAKOMLDPANYGTGMRWPRTPKBEAQSLEDLAGF 1720  
QY 503 -----PTTYYKAPAPAPAP-----TTPKET--APTTP-----KK 532  
Db 1721 KELFQTPDHTTESTDDKTTKIACKGPPPEBMDTPTSTRRRPKTPLGKRDIVBELSALKQ 1780  
QY 533 LTPPTT-----PEKAP-----TTPEKPAPTTP-----BELA----- 558  
Db 1781 LTQTHWDKVPGEDKGINVERETAKQKLDPAASVTCGSKQPRTPKGKAQPLEDLACLKE 1840  
QY 559 -----PTTPPEBPTTPPEPAPT----- 576  
Db 1841 LFTQPVCTDKPTTTHXETTKTIACRSQPDPVGVGTPTIFKPQSKRSLRKADVBEESIALRKRT 1900  
QY 577 -----TPKAAA-----PNTPKPEPA----- 590  
Db 1901 PSVGKAMDTPKPAAGGDEKDMKAMFGTPTVQKLDLGNLPGSKRWPQTPKEKAQALEDLAGF 1960  
QY 591 -----PTTPKEPAPTTPKEPA-----TTPKET----- 613  
Db 1961 KELFQTPGTDKPTTDEKTKIACKSPQDPDVPDTPASTKQRPKNLKNKADVEEBFLALRK 2020  
QY 614 -----APTTPK-----GTAPTTLKEPAPTTPPKAP----- 639  
Db 2021 TPSAGKAMDTPKPAVSDEKNINTFVETPVQKLDLGNLPGSKRQ--PQTPKEKALEDL 2078  
QY 640 ---KELAPTTPKEPTSTSDK-----PAP----- 660  
Db 2079 VGFKELFQTPGHTTEBMTDDKITEVSCSKSPQSPESFKTSRSKQRLKIPLVKMDKSEPLA 2138  
QY 661 -----TTPKGAPT-----TPK-----EPAPT-----TPKEPA----- 683  
Db 2139 VSKLTRTSGETTQHTTEPTGDSKSIKAFKESPKQILDPAASVTCGSRRLKTRKEKARALE 2198  
QY 684 -----PTTPKGAPTTLKEPAPTTPPKAPAPKEL 711  
Db 2199 DLVDFKELFSAPGHTBESMTIDKNTKIPCKSPPPPELTDATSTKRCF-KTRPRKEVKEEL 2257  
QY 712 AP-----TTTKGPTSTTSDEPA-----PTTPKE----- 734  
Db 2258 SAVERLUTQSGQSTHTEKHPASGDGKIKVLQKRAKKPNVVEEPPSRRRPRAPKAKAQL 2317  
QY 735 -----TAPTTPKEPAPTTPPKAPAPTTPPETPTTSEVSTPTTKE 774  
Db 2318 EDLAGFTELSETSGHTQESLTAGKATKIPCESPPLEVVDITASTKRLHTRVQKVQKEE 2377  
QY 775 PTHIHKSPDESTELSAEPTP-----KALENSPKPEPGVPTT----- 810  
Db 2378 PSAV-KFTQTSGETTDADKEPAGEDKGIKALKESAKQTPAPAASVTCGSRRRPRAPRESAQ 2436  
QY 811 -----KTPAA--TKPEMTTAKTKTERDLRTPE--TTTAAKPMKETAATTEKT 857  
Db 2437 AIEDLAGFKDPAAGHTEESMT---DDKTTKIPCKSSPELEDATSSKRRPRTAKQVEVK 2493  
QY 858 TE-----SKITATTTQVSTTQ-----DTTPPKITTLKTTLAPKVTTKKTIITTEIM 907  
Db 2494 EELLAVGKLTQTSGETTHDKPEPVGEGKGTAKF-----QPAKRNVDABDVI 2540  
QY 908 NKPEETAKPKDRA-----TNSKATTPKPKQKPTK-- 935  
Db 2541 GSRQRAPKPKAQLDLASFQELSQTGHTTELANGAADSFTSAPKQTPDSGKPLKIS 2600

QY 936 ---APK-KPT----- 941  
Db 2601 RRVLRAPKVEPVGDVVSTRDPVKSQKSNSTSLPPLFKRGGKSGSVGTGKRLRCMPAPE 2660  
QY 942 -----STTKPKTMRVRKPKTPTT---PRKMTSTMPBLNPTSRIAEAMLQTTTRPNQ- 990  
Db 2661 EIVEELPASKQKQVAPRARGKSSEPVVIMKRSURTSAKRIEPAEELNSNDMKTKEEHL 2720  
QY 991 ---TPNSKLVEVNPXKSDAGAGETPHMLL-----RPHFMEVPTFDMDYLP 1035  
Db 2721 QDSVPENKGISLRSRQDKTEAQOQITEVFLAERIEINRNEKKP---MKTSPEMD--- 2773  
QY 1036 RVENQGIINP 1046  
Db 2774 ---IONP 2777  
RESULT 14  
A48666  
cell proliferation antigen Ki-67, long form - human  
C:Species: Homo sapiens (man)  
C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 09-Jul-2004  
C:Accession: A48666  
R:Schlueter, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Gerdes  
J. Cell Biol. 123, 513-522, 1993  
A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, ubiq  
ins.  
A:Reference number: A48666; MUID:94043435; PMID:8227122  
A:Accession: A48666  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-3256 <SCH>  
A:Cross-references: UNIPROT:P46013; EMBL:X65550; MID:g415818; PIDN:CAA46519.1; PID:g4158  
C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat  
F:29-91/Domain: kinase interaction domain homology <KH>  
Query Match 15.0%; Score 834.9; DB 2; Length 3256;  
Best Local Similarity 15.9%; Pred. No. 2.5e-09;  
Matches 392; Conservative 138; Mismatches 420; Indels 1521; Gaps 94;  
QY 22 VSSQELSK-----GRCFSEFERGECDCD 46  
Db 741 ISNQKMDFKDLGSLGIAEMFKTPVKEQFQLTSTCHIAISNENLLGKQFGQTSGE----- 796  
QY 47 AQCKKYDKCCPDYSEF-----CAEEH-SVSE----- 71  
Db 797 -----PLLTSESFGNVFFSAQNAAKQPSKASPLRRQCIRENGNVAKTPRNTY 849  
QY 72 ---NORSSSSSSSSSTTIWKIKSSKNGSAANRELQ----- 105  
Db 850 KMTSLETKISDTETESKIVSTVNRSGRSTEFPRNIQKLPVESKSEETNTEIVCILKRGQ 909  
QY 106 -----KUKVON-----KKNRT-KKKPTPKPPVDEAG 132  
Db 910 KATLQORREGEKIEIRPPTYKENIELKENDKMKMRSRWTGCKAPMSDLTD--- 966  
QY 133 SGLDNGDFKVTTPDT-----STTQNKVSTSPKITAKPINRPSLPP--N 176  
Db 967 -----LKSPLDTELMDKTARGONLLQDQAKAPKSEKGIKTK--MPCQSLQPEPIN 1016  
QY 177 SDTS-----KETSITVKNKETTETKETTNNKQSTDKGE----- 211  
Db 1017 TPTHTTKQLKASLGKVGKVELLAVGKFTRT-SGETTHTHREPADGKSIKRTFKESPKI 1075  
QY 212 -----KTTSAKETQSIKTSADLAPTSKVLAKPTPKAE----- 245  
Db 1076 LDPAARVTGKMWPRTPKBEAQSLE-----DLAGFKELFQTPGSEESMTDEKTKIACK 1130  
QY 246 ---TTTK-----GPALTTPKEP----- 259  
Db 1131 SPPESVDPTSTKQWPKRSLRKADVEEBFLALRKLTSPSAGKAMLTPKPAGGDEKIKAF 1190  
QY 260 --TPT-----TPKBPAST-----TPKBP 275

```
Db 1191 MGTTPVQKLDLAGTLPQSKRQLQTPKEKAQALDLAGFKELFQTPGHTTELVAAAGKTKIP 1250
QY 276 -----TPTTIKSAPT-----TPKEPAP-----292
Db 1251 CDSQSPDPVDTPTSTKQPKRSIRKADVEGELLACRNLMPSAGKAMHTPK-----PSVGEK 1307
QY 293 -----TTTKSAPTTTKEPAPT-----308
Db 1308 DIIIFVGTVPQKLDLTENLTGSKRRRQTPKEBAQALDLTGFKELFQTPGHTTEVAAGK 1367
QY 309 TTKEPAPTTPKEPAPTTPKEPAPTTPKESAPTTPKE-----343
Db 1368 TTMPCESSPPESADT-----PTSTRQPKTFLEXRDVQKELSALKKLTQTSGETHTD 1421
QY 344 -----PATTPKPAPTTPKEPA-----361
Db 1422 KVPGGSEKINAPRETAQOKLDPAASVTGSKRHPT-KEKAQPLEDLAGMKELFQTPVCT 1480
QY 362 --PTTPK-----EPTPT-TPKEPAPTTPKE-----382
Db 1481 DXPTTHEKTTKACRSQDPVDTPTSSKQSKSLRKVDVEEFFALRKTPSAGKAMHT 1540
QY 383 PAPT-----TPKEPA-----392
Db 1541 PKPAVSGEKNIYAFMGTPVQKLDLTENLTGSKRRRLQTPKEKAQALDLAGFKELFQTRGH 1600
QY 393 -----PTAPKKA-----PTT 403
Db 1601 TEESMTNDKIAKVACKSQDPDLKDPASSKRRRLKTSGLGVGKVELLAVGLTQTSGET 1660
QY 404 PKEPAPT-----TPKEPA-----P 417
Db 1661 HHTEETGDKSMKAFMESPQKILDSASLTGSKRQLRTPKGSFVPEDLAGFIELFQTP 1720
QY 418 TTKEPSPTPK-----EP-----APTTSKAP-----440
Db 1721 SHTKE-SMTNEKTTKVSYRASQDVLDTPTSSKQPKRSRLKADTEEEFLAPRKLTFS 1779
QY 441 -----TTTKEPA-----447
Db 1780 KAMHTPKPVGEEKDINTFLGTPVQKLDQPGNLPGSNRRRLQTRKEKAQALELTGPRELF 1839
QY 448 -----PTK-----SAPTTKPESPTTTKE-----PA 470
Db 1840 QTPCTDNPTADEKTTKILCKSPQSDPAPTPTNKQPKRSRLKADVEEEFLAPRKLTFS 1899
QY 471 -----PTTPKEPAPT-----480
Db 1900 AGKAMHTPKAAVGEKDIINTFVGTTPVEKLDLGNLPGSKRRRQTPKEKAQALDLAGFKE 1959
QY 481 --TP-----KKPAPTTPKEPA-----494
Db 1960 LFGTPGHTBESMTDDKITEVCSKSPQDPVKTPTSSKQRLKISLGVGVKBEVLVPVGLT 2019
QY 495 -----PTTPKEPA-----502
Db 2020 QTSKGTQTHRETAGDKSIKAFKESAKQMLDPANYGTGMRWPRTPKKEAQSLEDLAGF 2079
QY 503 -----PTTPKPAPTAPKEPAP-----TTPKET-----APTTP-----KK 532
Db 2080 KELFQTPDHTESRTDDKTKIACKSPFPESMDTPTSTRRPKTPGKRDI VEELSALKQ 2139
QY 533 LPTTT-----PEKLAP-----TTPKEPAPTTP-----BELA-----558
Db 2140 LQQTHTDQVPGDEKGINVRETAKQKLDPAASVTGSKRQRTPKGAQPLEDLAGLKE 2199
QY 559 -----PTTPEEPTTPTEPAPT-----576
Db 2200 LFGTPVCTDKPTTHEKTTIACKSPQDPVGTPTIFKPOSKSLRKADVEEBSLAURKKT 2259
QY 577 -----TPKAAA-----PNTPKPEPA-----590
```

```
Db 2260 PSVGKAMDTPKPAGGDEKDMKAFMGTPVQKLDLGNLPGSKRWPTQTPKEKAQALEDLAGF 2319
QY 591 -----PTTPKEPAPTTPKEPAP-----TTPKET-----613
Db 2320 KELFQTPGTDKPTTDEKTKTKIACKSPQDPVDTPTASTKORPENLRKADVEEEFLALRKR 2379
QY 614 --APTTPK-----CTAPTTLKEPAPTTPKKEPAP-----639
Db 2380 TPSAGKAMDTPKPAVSDENKINFTVETPVQKLDLGNLPGSKRQ--PQTPKEKALEDL 2437
QY 640 -----KELAPTTTKEPTSTTSK-----PAP-----660
Db 2438 VGFKELFQTPGHTESMTDDKITEVCSKSPQESFKTSRSSKQRLKIPLVKVDMKKEPLA 2497
QY 661 --TTPKGTAPT-----TPK-----EPAPT-----TPKEPA-----683
Db 2498 VSKLRTSGETTQTHTEPTGDSIKAPKESPKQILDPAASVTGSSRRQLRTRKEKARALE 2557
QY 684 -----PTTPKGTAPTTLKEPAPTTPKPKAPKEL 711
Db 2558 DLVDFKELFSAPGHTESMTIDKNTKIPCKSPPELTDATSTKRCP-KTRPRKEVKEEL 2616
QY 712 AP-----TTTKGPTSTTSKPA-----PTTPKE-----734
Db 2617 SAVERLTQTSQGQSTHTEKPEASGDEGIKVLKQRAKKPNPVEEPPSRRRPRAPKEKAQPL 2676
QY 735 -----TAPTTPKEPAPTTPKPAPTTPPETPTPTTSEVGTPTTKE 774
Db 2677 EDLAGFTELSETSGHTQESLTAGKATKIPCESPPLEVVDTTASTKRLHTRVKOVKKEE 2736
QY 775 PTTIHKSPDESPELSABPTP-----KALENSKPEPGVPTT-----810
Db 2737 PSAV-KFTQTSGETTDADKEPAGEDGKI KALKESAKQTPAPAASVTGSSRRRRPRAPRESAQ 2795
QY 811 -----KTPAA--TKPEMTTTAKDKTTERDLRTPE--TTTAAFKMTKETAATTTEKT 857
Db 2796 ATEDLAGFKDPAAGHTEESMT--DDTKTKIPCKSSPELEDATSSKRRPRTEAOKVEVK 2852
QY 858 TE-----SKTATTTQVTSITQ-----DTPFKITTLTKTLAPKVTTTKTITITTEIM 907
Db 2853 EELLAVGKLTQTSGETTHTDKEPVGEGKGTAKF-----QPAKRNVDAEDVI 2899
QY 908 NKPEETAQPKDRA-----TNSKATTPKQKPTK--935
Db 2900 GSRQRPAPKEKAQPLEDLASFQELSQTGHTTEELANGAADSPSAPKOTPDGSKPLKIS 2959
QY 936 --APK-KPT-----941
Db 2960 RRVLRAPKPEVPGDVVSTRDPVKQSQKSNSTLPLPFKRGKGKGSVGTGKRLRCMPAPE 3019
QY 942 -----STKKPKTMPRVRKPKTTPT--PRKMTSTMPELNTPTSRIAEAMLOTTTTPNQ--990
Db 3020 EIVEELPASKQORVAPRARCKSEPVVIMKRSLSRTSAKIEPAEELNSDMKTNKEEHLK 3079
QY 991 --TNSKLVEVNPVKSSEDAGGAETPHMLL-----RPHVPEPTEDDMDYLP 1035
Db 3080 QDSVFNKGISRSRRQDKTEABQQITEVFVLAERIEINRNEKPP-----MKTSPEMD---3132
QY 1036 RVPNOGIIINP 1046
Db 3133 -----IQNP 3136
```

## RESULT 15

A41819

proline-rich peptides 637K precursor, prostatic - rat  
C:Species: Rattus norvegicus (Norway rat)

C:Date: 13-May-1992 #sequence\_revision 13-May-1992 #text\_change 09-Jul-2004

C:Accession: A42663; A41819; A31966; B20593; A20593

R:De Clercq, N.; Hemschoote, K.; Devos, A.; Peeters, B.; Heyns, W.; Rombauts, W.

J. Biol. Chem. 267, 9884-9894, 1992

A:Title: The 4.4-kilodalton proline-rich polypeptides of the rat ventral prostate are the exon.



```

QY 535 -----PTTPEKL-----APTPEKAPT-----PEELAPTP 562
Db 1921 WVPHIENTEKIVLHYAEPPTGPFVEPPELFLKTTKSRVQQTATMAASPKEMVSRAP 1980
QY 563 EE-----PTPTPEEPAP-----TPKAAAPNTPK-----EPA 590
Db 1981 ENKEAVLSGFGEDQSPSPNNMSLQDQELTSSQPHGWIHPHNTHGKIYLHYAEP- 2039
QY 591 PTPP-----KEPAPTPK----- 603
Db 2040 PTGPFVEPPELFLKTTKSKPVEWTLTRDKSRKEMVSPKPYEEBAVLPHVHGQESRS 2099
QY 604 -----EPAPTPP-----KETAP 615
Db 2100 PPNMSLQSLQELTSSQPHGVPHPNTHGKIYLHYAEP-PTGPFVEPPELFLKTTKS 2158
QY 616 TTPKGTAPTTLKEP----- 629
Db 2159 KPVGQTATRMVKSPEEMVSLDPENKEAVFPAQEGKGESPPNMSLQSLDHLELFWSQP 2218
QY 630 ----APTTPKKP-----APKELAPT 645
Db 2219 HGWIHPHNTPKIYLHYAEPPTGPFVEPPELFLKTTKSPVQGTTEMMAKSPKEMVSO 2278
QY 646 TTK----- 648
Db 2279 TPYEKAVLSGPGEDQSPSPNTSLKSLDQEVAMSSQPHGVPHPKTPGKIYLSIE 2338
QY 649 -----EPT-----STTSDKAPATPK----- 664
Db 2339 PPPGPFVKPTDLILVTKTKSPAETPRRIDKLLKEMVPHSPYEBAVFPAGHGEQDES 2398
QY 665 GTAPTTPKEPA-----PTTP-----KEPAPTPKGTAP-----TT 694
Db 2399 GSPNNPLQDQELTSSQPHGVPHHNTPGKIYLHYAEP-PTGPFVEPPELFLKTT 2457
QY 695 LKEPAPTTPKPA--PKELA----- 712
Db 2458 KSKPVQSGPKIAKSPKEMVSTPYEKAVLSGPGEDQSPSPNNMSLSLDQEVMS 2517
QY 713 -----PTTKG-----PT-----STTSDKAPATTP-----KET 735
Db 2518 QPHSGVPHPKTPGKIYLSHISBPPPGPFVKPTDLILVTKTKSPAETPRRIDKLLKEM 2577
QY 736 APTTPE-----PA-----PTTPKKP-----APTTPETP----- 759
Db 2578 VPHSPEYEEAVFPAGHGEQDESPPNNPLQDQELTSSQPHGVPHHNTPGKIYLH 2637
QY 760 ----PTTTSVSTP-----TTTKE-----PTTHKSPDE----- 784
Db 2638 YAEPTGPFVEPPELFLKTTKSPKVOGTQMAKSPPEMVSLSPKNETVFPAGKQGD 2697
QY 785 -----STPELS-----AP----- 793
Db 2698 EGSPPNNISLQSLDQELTSSQPHGWIHPHNTHGKIYLHYAEPPTGPFVEPPELFLKTT 2757
QY 794 -----TPKALENSPKE----- 804
Db 2758 TSKSPMQSGPROIDKSPKEMFTQSPYEESLIPAHABGOESRAPPHFSLOPLDQELSLS 2817
QY 805 -----PGVP-----TTKPAATKP-EMTTAKDXT 828
Db 2818 SHPHGWIHPHNTPKIYLHYAEPPTGPFVEPPELFLKTTK-----SKPVQGTATKTDKS 2873
QY 829 TERDLTTPETTTAA-----PKMT-----KETATTTT----- 855
Db 2874 PEDRVSTPYEKAVLSGPGEDQSPSPNNMSLQSLDQELALISSQPHGWIHPHNPADK 2933
QY 856 -----KTESK-ITATTQVSTTQ----- 875
Db 2934 IYLHYAEPPTGPFVEPPELFLKTTKSPKVOGTQMAKSPKEMVSTPYEKAADLSAFG 2993

```

```

QY 876 -----DTTPE-----KIT 893
Db 2994 ENRDESPSSNNMSLHLDQELSSQPHGWIHPHNTHGKIYLHYAEPPTGPFVEPPELFL 3053
QY 884 TLKTTTTLAPKVTTTKTITTTIMNKPEETAK-PKD----- 918
Db 3054 FLKTTTKSKP-----VOGSPSQUIDKSPKEVFSQSPSESVLPAQAGQESR 3100
QY 919 -----RATNS 923
Db 3101 APPHMSLQDQKLTLSSHPHGWIHPHNTPGKIYLHYAEPPTGPFVEPPELFLKTTKS 3160
QY 924 KATTPKPKPTKAPK----- 939
Db 3161 KPVOGSPROVDKSPKEMFTQSPYEESVLPAAQAGQESRAPPHMSLQDQDLTLSSHP 3220
QY 940 -----PTSTKK-----PKTMRVRKP-----KTTPT-----PRMTST 967
Db 3221 HGWIHPHNTPKIYLHYAEPPTGPFVEPPELFLKTTKSKPAQWTPQTQMAKSPPEMVSL 3280
QY 968 MPE-----LNPTSRIAEAMLQTT-----RPN----- 989
Db 3281 SPENKETVFPAGKQGESISPPHMSLQDQDLTPSSHHPGWIPHPHNTHGKIYLHYAE 3340
QY 990 -----OTPNKLVENP----- 1001
Db 3341 PTGPFVEPPELFLKTTKSPVQSPROIDKSPKEVFTQSPYEESVLPAAQAGQESR 3400
QY 1002 -----KS 1003
Db 3401 APPHMSLQDQDLTLSSHPHGWIHPHNTHGKIYLHYAEPPTGPFVEPPELFLKTTKS 3460
QY 1004 EDAGG-----AEGE-----TPHMLLR-----P 1020
Db 3461 KPVGSPROIDKSHKEMFTQSPYEESVLPAAQAGQESRAPPHMSLQDQDLTLSSHP 3520
QY 1021 HVFMPE--VTPDMVDYLPV-PNOGIIINP 1046
Db 3521 HGWIHPHNTPKIYLHYAEPPTGPFVEP 3549

```

Search completed: October 13, 2004, 11:58:55

Job time : 65.3671 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 81.1065 Seconds  
(without alignments)  
4449.477 Million cell updates/sec

Title: SEQ1-E

Perfect score: 5373

Sequence: 1 MAWKTLPIVILLLSVFVIQ.....DMDVLRVPVQGIINPMLS 1006

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.1

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_23Sep04:.\*  
1: Geneseq1980s:.\*  
2: Geneseq1990s:.\*  
3: Geneseq2000s:.\*  
4: Geneseq2001s:.\*  
5: Geneseq2002s:.\*  
6: Geneseq2003as:.\*  
7: Geneseq2003bs:.\*  
8: Geneseq2004s:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	5373	100.0	1270	8	Adk67911 Human ext
2	5358.9	99.7	1311	8	Adk67912 Human ext
3	5349	99.6	1320	7	Adk65819 Angiogene
4	5339.6	99.4	1299	4	Aam24322 Human EST
5	5339.6	99.4	1404	2	AAR26049
6	5339.6	99.4	1404	4	AAB29773
7	5339.6	99.4	1404	4	AAB60568 Human meg
8	5339.6	99.4	1404	8	Adm98014 Human meg
9	5339.6	99.4	1415	4	Aau32262 Novel hum
10	5330.6	99.2	1404	7	Adk65839 Angiogene
11	2920	54.3	546	4	ABU53252
12	2888	53.8	551	4	ABU53253
13	2880.1	53.6	902	4	AAB29778 Human tes
14	2757	51.3	513	4	ABU53254 Human tes
15	1694.8	31.5	452	2	AAR80041 Human meg
16	1409.9	26.2	538	5	AAO18834 3' cartil
17	1335.3	24.9	5179	4	AAM24516 C899P pre
18	1335.3	24.9	5179	6	ABP55365 Human col
19	1335.3	24.9	5179	6	ABO07258 Human p53
20	1335.3	24.9	5179	7	ADD48091 Human Pro
21	1335.3	24.9	5179	7	ADD44998 Human Pro
22	1146.1	21.3	292	5	Aau11261 Human HAP
23	1108.6	20.6	1664	2	AAW43106 C. thermo
24	1099.6	20.5	717	4	ABU53144 Human tes
25	1097	20.4	214	4	ABU53255 Human tes

## ALIGNMENTS

## RESULT 1

ADK67911  
ID ADK67911 standard; protein; 1270 AA.

XX AC ADK67911;

XX DT 06-MAY-2004 (first entry)

XX DE Human extracellular messenger (EXMES) polypeptide.

XX KW Human; extracellular messenger; EXMES; respiratory-gen.; anti-allergic;  
KW antiasthmatic; antiinflammatory; antidiabetic; neuroprotective;  
KW muscular-gen.; antiarthritic; osteopathic; hepatotropic; antiparasitic;  
KW virucide; fungicide; antiparasitic; protozoacide; antihelminthic;  
XX cytotstatic; gene therapy.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers  
XX FT Peptide 1..29

XX FT /label= Signal peptide  
XX FT /note= "Spans residues 1 to 18, 20, 21, 24, 27 or 29 according to identification method"

XX FT WQ2004013292-A2.

XX PD 12-FEB-2004.

XX PF 30-JUL-2003; 2003WO-US024084.

XX PR 02-AUG-2002; 2002US-0400810P.

XX PR 19-SEP-2002; 2002US-0412197P.

XX PR 04-OCT-2002; 2002US-0416004P.

XX PR 08-NOV-2002; 2002US-0424862P.

XX (INCY-) INCYTE CORP.

XX PI Elliott VS, Khare R, Tran UK, Swarnakar A, Marquis JP;

XX PI Richardson TW, Emerling BM, Lindquist BA, Chawla NK, Ramkumar J;  
XX PI Lee SY;

XX DR WPI: 2004-157116/15.

XX DR N-PSDB; ADK67916.

XX PT New extracellular messengers and nucleic acids, useful for diagnosing,  
PT treating or preventing e.g. allergies, anemia, Crohn's disease, diabetes  
PT mellitus, myasthenia gravis, osteoporosis, psoriasis, cancers or  
PT autoimmune thyroiditis.

26 1075.7 20.0 8991 6 ABU08487  
27 1047.1 19.5 763 3 AAG38942  
28 1041.7 19.4 1795 4 ABB69806  
29 1039.7 19.4 1049 4 ABB61364  
30 1008.8 18.8 2284 4 ABB71434  
31 987.9 18.4 770 4 ABUS3141  
32 981 18.3 778 4 ABUS3143  
33 974.4 18.1 692 4 ABUS3155  
34 964.1 17.9 717 4 ABUS3150  
35 964.1 17.9 717 4 ABUS3149  
36 964.1 17.9 717 4 ABUS3151  
37 964.1 17.9 717 4 ABUS3145  
38 964.1 17.9 717 4 ABUS3148  
39 964.1 17.9 717 4 ABUS3147  
40 964.1 17.9 717 4 ABUS3146  
41 955.4 17.8 5703 8 ADL23265  
42 952.1 17.7 745 4 ADL23265  
43 950.4 17.7 4315 5 ABP43908  
44 948.6 17.7 2112 4 ABB60403  
45 932.7 17.4 1538 7 ADI21202

Abu08487 S. pneumo  
Aag38942 Arabidops  
Abb69806 Drosophil  
Abb61364 Drosophil  
Abb71434 Drosophil  
Abu53141 Human tes  
Abu53143 Human tes  
Abu53155 Human tes  
Abu53150 Human tes  
Abu53149 Human tes  
Abu53151 Human tes  
Abu53145 Human tes  
Abu53148 Human tes  
Abu53147 Human tes  
Abu53146 Human tes  
Adl23265 Human MUC  
Adl23265 Human MUC  
Abu53142 Human tes  
Abp43908 MUC5B par  
Abb60403 Drosophil  
Adi21202 Novel hum

XX PS Claim 59; SEQ ID NO 4; 165pp; English.

XX CC The present sequence is that of novel human extracellular messenger

CC (EXMES) Incyte ID NO: 7513017CDI polypeptide. The protein shows homology

CC to human megakaryocyte stimulating factor. The invention provides EXMES

CC polynucleotides and polypeptides, as well as expression vectors, host

CC cells, antibodies, agonists and antagonists, and methods for diagnosing,

CC treating or preventing disorders associated with aberrant expression of

CC EXMES, especially autoimmune and inflammatory disorders, cell

CC proliferative disorders and endocrine disorders, e.g. adult respiratory

CC distress syndrome, allergy, anaemia, erythroblastosis foetalis, Crohn's

CC disease, diabetes mellitus, myasthenia gravis, osteoarthritis,

CC osteoporosis, pancreatitis, psoriasis, viral, bacterial, fungal,

CC parasitic, protozoal or helminthic infections, cancers, autoimmune

CC thyroiditis, cretinism, Plummer's disease or thyroid carcinoma.

CC Embodiments also provide methods for using the purified EXMES and/or

CC their encoding polynucleotides for facilitating the drug discovery

CC process, including determining of efficacy, dosage, toxicity and

CC pharmacology, and for investigating the pathogenesis of diseases and

CC medical conditions.

XX SQ Sequence 1270 AA;

Query Match 100.0%; Score 5373; DB 8; Length 1270;

Best Local Similarity 100.0%; Pred. No. 2.4e-142;

Matches 1006; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQELSCKGRGCFESFERGECDCDAQCKKYDKCCPDYE 60

DB 1 MAWKTLPIYLLLLSVFVIQVSSQELSCKGRGCFESFERGECDCDAQCKKYDKCCPDYE 60

QY 61 SFCAEVKDNKNTKKTKPKPPVVDAGSLDNGDFKVTPTDSTTQHNVKSTSPKITT 120

DB 61 SFCAEVKDNKNTKKTKPKPPVVDAGSLDNGDFKVTPTDSTTQHNVKSTSPKITT 120

QY 121 AKENRPSLPNNSDTSKETSITVNEKTTVETTTNNKQSTDDGKEKITSKEQTOSIE 180

DB 121 AKENRPSLPNNSDTSKETSITVNEKTTVETTTNNKQSTDDGKEKITSKEQTOSIE 180

QY 181 KTSAKOLAPTSKVLAKTPKAEITTKGPAITTKPEPTTTPKPEPASTPKPEPTTIKSA 240

DB 181 KTSAKOLAPTSKVLAKTPKAEITTKGPAITTKPEPTTTPKPEPASTPKPEPTTIKSA 240

QY 241 PTTKPEPAPTTTKSAPTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKSAPTTKPE 300

DB 241 PTTKPEPAPTTTKSAPTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKSAPTTKPE 300

QY 301 PAPTTPKKPAPTTPKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTT 360

DB 301 PAPTTPKKPAPTTPKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTT 360

QY 361 PKEPAPTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTK 420

DB 361 PKEPAPTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTK 420

QY 421 TTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPE 480

DB 421 TTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPE 480

QY 481 ETAPTTPKLPTTPPEKLAFTTPKPEPAPTTPEELAPTTPEEPTTTPPEEAPTTPKAAAP 540

DB 481 ETAPTTPKLPTTPPEKLAFTTPKPEPAPTTPEELAPTTPEEPTTTPPEEAPTTPKAAAP 540

QY 541 NTPKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPE 600

DB 541 NTPKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPE 600

QY 601 PTTTKPESTTSKPEAPTTKGTAPTTKPEPAPTTKPEPAPTTKGTAPTTLKPEPAPTT 660

DB 601 PTTTKPESTTSKPEAPTTKGTAPTTKPEPAPTTKPEPAPTTKGTAPTTLKPEPAPTT 660

QY 661 KKPAPKELAPTTTKGPTSTTSKAPATTTPKETAFTTPKPEPAPTTPKKPAFTTTPPTT 720

DB 661 KKPAPKELAPTTTKGPTSTTSKAPATTTPKETAFTTPKPEPAPTTPKKPAFTTTPPTT 720

QY 721 SEVSTPTTTTKEPTTIHKSPDESTPELSAEPTPKALENSKPEGVPPTTKTPAATKPEMTT 780

DB 721 SEVSTPTTTTKEPTTIHKSPDESTPELSAEPTPKALENSKPEGVPPTTKTPAATKPEMTT 780

QY 781 AKDKTTERDLRTTPETTTAAPKWKETATTBKTTESKITATTQVTSSTTQDTPPKIT 840

DB 781 AKDKTTERDLRTTPETTTAAPKWKETATTBKTTESKITATTQVTSSTTQDTPPKIT 840

QY 841 TLKTTTLAPKVTITTKKTIITTEIMNKPEETAPEAKPKDRATNSKATTPKPEQPTKAPKPTST 900

DB 841 TLKTTTLAPKVTITTKKTIITTEIMNKPEETAPEAKPKDRATNSKATTPKPEQPTKAPKPTST 900

QY 901 KKPKNMVRVRKPKTTPTPKAMTSTMPELNPTSRIAEAMLTOTTRPNQOTENSKLVEVNPKS 960

DB 901 KKPKNMVRVRKPKTTPTPKAMTSTMPELNPTSRIAEAMLTOTTRPNQOTENSKLVEVNPKS 960

QY 961 EDAGGAGETPHMLLRPHVFMPEVTPDMYLRPVPNQGIINPMLS 1006

DB 961 EDAGGAGETPHMLLRPHVFMPEVTPDMYLRPVPNQGIINPMLS 1006

RESULT 2

ADK67912

ID ADK67912 standard; protein; 1311 AA.

XX AC ADK67912;

XX DT 06-MAY-2004 (first entry)

XX DE Human extracellular messenger (EXMES) polypeptide.

XX KW Human; extracellular messenger; EXMES; respiratory-gen.; anti-allergic;

KW antiasthmatic; anti-inflammatory; antidiabetic; neuroprotective;

KW muscular-gen.; antiarthritic; osteopathic; hepatotropic; antipsoriatic;

KW virucide; fungicide; antiparasitic; protozoacide; antihelminthic;

KW cytostatic; gene therapy.

XX OS Homo sapiens.

XX FH Key

XX FT Peptide

XX FT 1. .30

XX FT /label= Signal\_peptide

XX FT /notes= "Spans residues 1 to 18, 20, 21, 24, 29 or 30

XX FT according to identification method"

XX WO2004013292-A2.

XX PN 12-FEB-2004.

XX PD 30-JUL-2003; 2003WO-US024084.

XX PF 02-AUG-2002; 2002US-0400810P.

XX PR 19-SEP-2002; 2002US-0412197P.

XX PR 04-OCT-2002; 2002US-0416004P.

XX PR 08-NOV-2002; 2002US-0424862P.

XX PA (INCY-) INCYTE CORP.

XX XX Elliott VS, Khare R, Tran UK, Swarnakar A, Marquis JP;

PI Richardson JW, Emerling BM, Lindquist EA, Chawla NK, Ramkumar J;

PI Lee SY;

XX WPI; 2004-157116/15.

DR N-PSDB; ADK67917.

XX New extracellular messengers and nucleic acids, useful for diagnosing,

PT treating or preventing e.g. allergies, anemia, Crohn's disease, diabetes

PT mellitus, myasthenia gravis, osteoporosis, psoriasis, cancers or

PT autoimmune thyroiditis.

XX



PS	Claim 60; SEQ ID NO 5; 165pp; English.	
XX	The present sequence is that of novel human extracellular messenger (EXMES) Incyte ID NO: 7513018CD1 polypeptide. The protein shows homology to human megakaryocyte stimulating factor. The invention provides EXMES polynucleotides and polypeptides, as well as expression vectors, host cells, antibodies, agonists and antagonists, and methods for diagnosing, treating or preventing disorders associated with aberrant expression of EXMES, especially autoimmune and inflammatory disorders, cell proliferative disorders and endocrine disorders, e.g. adult respiratory distress syndrome, allergy, anaemia, erythroblastosis foetalis, Crohn's disease, diabetes mellitus, myasthenia gravis, osteoarthritis, osteoporosis, pancreatitis, psoriasis, viral, bacterial, fungal, parasitic, protozoal or helminthic infections, cancers, autoimmune thyroiditis, cretinism, Plummer's disease or thyroid carcinoma. Embodiments also provide methods for using the purified EXMES and/or their encoding polynucleotides for facilitating the drug discovery process, including determining of efficacy, dosage, toxicity and pharmacology, and for investigating the pathogenesis of diseases and medical conditions.	
XX	Sequence 1311 AA;	
XX	Query Match 99.7%; Score 5358.9; DB 8; Length 1311; Best Local Similarity 96.1%; Pred. No. 6.2e-142; Matches 1006; Conservative 0; Mismatches 0; Indels 41; Gaps 1;	
QY	1 MAWKTLPIYLLLSLVFV1QOVSSQ----- 25	
DB	1 MAWKTLPIYLLLSLVFV1QOVSSQDLSSCAGRGYSRDATCNCDYNGHYMECCPDF 60	
QY	26 -----ELSCGRCFESFERGECDCDAQCCKYDKCPDYSEFCAEVKONKNRKKPT 79	
DB	61 KRVCIAELSCGRCFESFERGECDCDAQCCKYDKCPDYSEFCAEVKONKNRKKPT 120	
QY	80 PRPPVDEAGSLDNGDFKVTTPDSTTQHNKYSTSPKITTAKPINRPSLPNSDTSKE 139	
DB	121 PRPPVDEAGSLDNGDFKVTTPDSTTQHNKYSTSPKITTAKPINRPSLPNSDTSKE 180	
QY	140 TSLTVNKEITVEKETTNTKQSTDPGKEKTSKAKETQSIKTSKOLAPTSKVLAKPTP 199	
DB	181 TSLTVNKEITVEKETTNTKQSTDPGKEKTSKAKETQSIKTSKOLAPTSKVLAKPTP 240	
QY	200 KAETTTKGALTPKBPPTTPKPEASTTPKPEPTTIKSAPTTPKPEAPTTTKSAPTTP 259	
DB	241 KAETTTKGALTPKBPPTTPKPEASTTPKPEPTTIKSAPTTPKPEAPTTTKSAPTTP 300	
QY	260 KEPAPTTKPEAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP 319	
DB	301 KEPAPTTKPEAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP 360	
QY	320 TTPKPEPTTPKPEAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTK 379	
DB	361 TTPKPEPTTPKPEAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTK 420	
QY	380 PSPTTKPEAPTTTKSAPTTTKSAPTTTKSAPTTTKSAPTTTKSAPTTTKSAPTTP 439	
DB	421 PSPTTKPEAPTTTKSAPTTTKSAPTTTKSAPTTTKSAPTTTKSAPTTTKSAPTTP 480	
QY	440 KKEAPTTKPEAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP 499	
DB	481 KKEAPTTKPEAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP 540	
QY	500 PTTPEKAPTTPEELAPTTPEEPPTTPPEEPAPTTKBPAPTTKBPAPTTKBPAPTT 559	
DB	541 PTTPEKAPTTPEELAPTTPEEPPTTPPEEPAPTTKBPAPTTKBPAPTTKBPAPTT 600	
QY	560 KEPAPTTKPEAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP 619	
DB	601 KEPAPTTKPEAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP 660	
QY	620 PKGTAPTTKPEAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP 679	
DB	679 PKGTAPTTKPEAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP 720	
DB	680 TSDKAPTTKPEAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP 739	
DB	721 TSDKAPTTKPEAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP 780	
QY	740 DESTPELSASPTPKALENSPKESGVPTTKPAATKPEMTTAKDKTTERDLRTPETTTA 799	
DB	781 DESTPELSASPTPKALENSPKESGVPTTKPAATKPEMTTAKDKTTERDLRTPETTTA 840	
QY	800 APKMTKETATTTKTTESKITATTTQVTSSTTQDTPFKITTLTKTLAPKVTTKKTIIT 859	
DB	841 APKMTKETATTTKTTESKITATTTQVTSSTTQDTPFKITTLTKTLAPKVTTKKTIIT 900	
QY	860 TTIMNKPSTAKPKORATNSKATTPKQKPTKAPKPTSTKPKTMPRVKPKTTPPTPR 919	
DB	901 TTIMNKPSTAKPKORATNSKATTPKQKPTKAPKPTSTKPKTMPRVKPKTTPPTPR 960	
QY	920 KMTSTMPCLNPTSRIAEAMLOTTTRNQTPNSKLVENPKSEADAGGAGETPHMLLRPHV 979	
DB	961 KMTSTMPCLNPTSRIAEAMLOTTTRNQTPNSKLVENPKSEADAGGAGETPHMLLRPHV 1020	
QY	980 FMEVTPDMDYLPVNPVNOGIIINPMLS 1006	
DB	1021 FMEVTPDMDYLPVNPVNOGIIINPMLS 1047	
RESULT 3		
ID	ADK65819	
AC	ADK65819 standard; protein; 1320 AA.	
XX	ADK65819;	
XX	06-MAY-2004 (first entry)	
XX	Angiogenesis-differentially expressed protein ANH0316.	
DE	cytostatic; cardiant; vasotropic; antiarteriosclerotic;	
XX	angiogenesis inhibitor; angiogenesis stimulator; angiogenic index;	
KW	gene expression; cancer; coronary artery disease; myocardial ischemia;	
KW	coronary arteriosclerosis; forensic medicine.	
XX	Homo sapiens.	
OS	WO2003066831-A2.	
XX	14-AUG-2003.	
PD	07-FEB-2003; 2003WO-US003848.	
XX	07-FEB-2002; 2002US-00067482.	
PR	10-JUN-2002; 2002US-00164595.	
PR	16-AUG-2002; 2002US-0403649P.	
PR	03-JAN-2003; 2003US-0437746P.	
XX	(ORIG-) ORIGENE TECHNOLOGIES INC.	
PA	Sun Z, Li X, Kovacs KF, Fan W, Jay G;	
XX	WPI; 2003-731502/69.	
PI	N-PSDB; ADK65818.	
DR	Determining the angiogenic index of a tissue or cell sample using	
XX	expression levels of differentially expressed genes, useful for	
PT	diagnosing or treating cancer, coronary artery disease, myocardial	
PT	ischemia and/or arteriosclerosis.	
XX	Claim 23; SEQ ID NO 58; 296pp; English.	
PS	The invention relates to a method of determining the angiogenic index of	
XX	a tissue or cell sample comprising assessing, in a sample, the expression	
CC	levels of one or more differentially-expressed gene from any of 34 DNA	
CC	sequences, given in the specification, where the levels are indicative of	

CC the angiogenic index. The methods and compositions of the present  
CC invention are useful for diagnosing, preventing and/or treating cancer,  
CC coronary artery disease, myocardial ischemia or coronary  
CC arteriosclerosis. They can also be used in research, drug discovery and  
CC forensic medicine involving angiogenesis. This sequence corresponds to  
CC one of the differentially expressed proteins of the invention.  
XX  
SQ Sequence 1320 AA;  
Query Match 99.6%; Score 5349; DB 7; Length 1320;  
Best Local Similarity 95.1%; Pred. No. 1.2e-141;  
Matches 1004; Conservative 0; Mismatches 2; Indels 50; Gaps 1;  
QY 1 MAWKTLPIYLLLLSVFVIQOVSSQELSKGRCFESFERGRCDCDAQCKKYDKCCPDYE 60  
DB 1 MAWKTLPIYLLLLSVFVIQOVSSQELSKGRCFESFERGRCDCDAQCKKYDKCCPDYE 60  
QY 61 SFCA-----EYKDNK 70  
DB 61 SFCAEHNTPSPSSKKARPPSGASQTIKSTYKSPKPNKKTKKVISEEITEVKDNK 120  
QY 71 KNRKKKPTPKPPVDEAGSLDNGDFKVTTPDTSTQHNKVSTSPKITTAKPINRPSL 130  
DB 121 KNRKKKPTPKPPVDEAGSLDNGDFKVTTPDTSTQHNKVSTSPKITTAKPINRPSL 180  
QY 131 PPNDSISKETSLVNEKTTVETKTTTNKQTSKGKETSKEKTSKAKDLAPT 190  
DB 181 PPNDSISKETSLVNEKTTVETKTTTNKQTSKGKETSKEKTSKAKDLAPT 240  
QY 191 SKVLAKPTKAEITTKGPAITTKPEPTTPKPEASTTKPEPTTIKSAPTTPKEPAPT 250  
DB 241 SKVLAKPTKAEITTKGPAITTKPEPTTPKPEASTTKPEPTTIKSAPTTPKEPAPT 300  
QY 251 TTKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 310  
DB 301 TTKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 360  
QY 311 PTPKEPAPTTPKEPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 370  
DB 361 PTPKEPAPTTPKEPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 420  
QY 371 EPAPTTTKESPPTPKPEPAPTTTKSAPTTKPEPAPTTPKSAPTTPKEPAPTTPKEPAPT 430  
DB 421 EPAPTTTKESPPTPKPEPAPTTTKSAPTTKPEPAPTTPKSAPTTPKEPAPTTPKEPAPT 480  
QY 431 PKPEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 490  
DB 481 PKPEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 540  
QY 491 TPTTPEKLAPTPKPEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 550  
DB 541 TPTTPEKLAPTPKPEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 600  
QY 551 PKPEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 610  
DB 601 PKPEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 660  
QY 611 TSDKPAPTTPKGATPTPKPEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 670  
DB 661 TCDKPAPTTPKGATPTPKPEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 720  
QY 671 TTTKGPSTTSDKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 730  
DB 721 TTTKGPSTTSDKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 780  
QY 731 EPTTIHKSPDESTPELSAETPKALENSPKPEPGVPTTKTAAKPEMTTTAKDKTTERDL 790  
DB 781 EPTTIHKSPDESTPELSAETPKALENSPKPEPGVPTTKTAAKPEMTTTAKDKTTERDL 840  
QY 791 RTTPETTTAAPKMTKETATTTTEKTESKITHATTTQVTSITTTQDTPPKITLTKTTLAPK 850  
DB 841 RTTPETTTAAPKMTKETATTTTEKTESKITHATTTQVTSITTTQDTPPKITLTKTTLAPK 900

QY 851 VTTTKTKTTTITIMNKPEETAKPKDRATNSKATTPKQKPTKAPKPTSTKKPKXTMPVR 910  
DB 901 VTTTKTKTTTITIMNKPEETAKPKDRATNSKATTPKQKPTKAPKPTSTKKPKXTMPVR 960  
QY 911 KPKTTTPPKMTSTMPELNPTSRIAEAMLTQTTTRNQTSNSKLVEVNPKESEDAGGAEGT 970  
DB 961 KPKTTTPPKMTSTMPELNPTSRIAEAMLTQTTTRNQTSNSKLVEVNPKESEDAGGAEGT 1020  
QY 971 PHMLLRPHVFMPEVTPDMDYLRVNPNOGIIINPMLS 1006  
DB 1021 PHMLLRPHVFMPEVTPDMDYLRVNPNOGIIINPMLS 1056  
RESULT 4  
ID AM24322 standard; protein; 1299 AA.  
XX AC  
XX AM24322;  
XX 12-OCT-2001 (first entry)  
XX Human EST encoded protein SEQ ID NO: 1847.  
DE Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;  
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;  
KW diagnostics; forensic test; gene mapping; genetic disorder; biodiversity;  
KW gene therapy; nutrition.  
XX Homo sapiens.  
XX WO200154477-A2.  
XX 02-AUG-2001.  
XX 25-JAN-2001; 2001WO-US002687.  
XX 25-JAN-2000; 2000US-00491404.  
PR 17-JUL-2000; 2000US-00617746.  
PR 03-AUG-2000; 2000US-00631451.  
PR 15-SEP-2000; 2000US-00663870.  
XX (HYSE-) HYSEQ INC.  
XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;  
PI Cao Y, Drmanac RA, Zhang J, Wehrman T;  
XX WPI; 2001-476164/51.  
DR N-PSDB; AAH98981.  
XX Isolated polypeptide for treatment of diseases, diagnostics, raising  
PT antibodies and research use.  
XX Claim 20; Page 1198-1201; 1275pp; English.  
XX The present invention provides the protein and coding sequences of novel  
CC proteins from a variety of organisms, including human, dog, cat, horse,  
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea  
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)  
CC from the organism of interest. They can be used in diagnostics,  
CC forensics, gene mapping, identification of mutations, to assess  
CC biodiversity and for nutritional purposes. The present sequence is a  
CC protein of the invention  
XX  
SQ Sequence 1299 AA;  
Query Match 99.4%; Score 5339.6; DB 4; Length 1299;  
Best Local Similarity 88.2%; Pred. No. 2.1e-141;  
Matches 1006; Conservative 0; Mismatches 0; Indels 134; Gaps 2;  
QY 1 MAWKTLPIYLLLLSVFVIQOVSSQ----- 25  
DB 1 MAWKTLPIYLLLLSVFVIQOVSSQDLSSCAGRCGYSRATCNCYDNCQHYMECCPDF 60

```
QY 26 -----ELSKGRCFESFERGECDAQCKYDKCCPDYESFCAE----- 65
Db 61 KRVTAELSCKGRCFESFERGECDAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
QY 66 ----- 65
Db 121 PPSGASQIKSTTKRSPKPNKXKTKVIESEEITEHSHVSENOESSSSSSSSSTIWI 180
QY 66 -----VKDNKKNTKKKPTPKPPVWDEAGSLDNGDFKVTTPDTST 106
Db 181 KIKSSKNSAANRELQKKLVKDNKNKTKKPTPKPPVWDEAGSLDNGDFKVTTPDTST 240
QY 107 TOHNKUSTSKITTAKEINPRPSLPNSDTSKETSITVNKETTIVETKETTITNNKQTSIDG 166
Db 241 TOHNKUSTSPKITTAKEINPRPSLPNSDTSKETSITVNKETTIVETKETTITNNKQTSIDG 300
QY 167 KEKTSKAKETQSIKETSASAKOLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTTPKEPAS 226
Db 301 KEKTSKAKETQSIKETSASAKOLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTTPKEPAS 360
QY 227 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKKEP 286
Db 361 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKKEP 420
QY 287 APTTTKSAPTTPKEPAPTTTKGAAPTTPKEPAPTTTPKEPTTTPKEPAPTTTKGPATTPK 346
Db 421 APTTTKSAPTTPKEPAPTTTKGAAPTTPKEPAPTTTPKEPTTTPKEPAPTTTKGPATTPK 480
QY 347 EPAPTAPKDPAPTTPKEPAPTTTPKEPAPTTTKPSPTTPKEPAPTTTKSAPTTPKEPAPT 406
Db 481 EPAPTAPKDPAPTTPKEPAPTTTPKEPAPTTTKPSPTTPKEPAPTTTKSAPTTPKEPAPT 540
QY 407 TTKSAPTTPKEPSPTTTPKEPAPTTTPKEPAPTTTPKAPAPTTPKEPAPTTTKKPTTKK 466
Db 541 TTKSAPTTPKEPSPTTTPKEPAPTTTPKEPAPTTTPKAPAPTTPKEPAPTTTKKPTTKK 600
QY 467 APTAPKEPAPTTPKETAPTTPKKLTPTPKLAAPTTPPEKAPATTPEELAPTTPEEPTPTT 526
Db 601 APTAPKEPAPTTPKETAPTTPKKLTPTPKLAAPTTPPEKAPATTPEELAPTTPEEPTPTT 660
QY 527 PEPAAPTTPKAAAPNTPKAPAPTTPKEPAPTTTPKEPAPTTPKETAAPTTPKGATPATTILKEP 586
Db 661 PEPAAPTTPKAAAPNTPKAPAPTTPKEPAPTTTPKEPAPTTTPKETAPTTPKGATPATTILKEP 720
QY 587 APTTPKAPAPKELAPTTTKEPTSTTSKAPAPTTPKGTAPTTPKEPAPTTTPKEPAPTTPKG 646
Db 721 APTTPKAPAPKELAPTTTKEPTSTTSKAPAPTTPKGTAPTTPKEPAPTTTPKEPAPTTPKG 780
QY 647 TAPTTLKEPAPTTPKKAPKELAPTTTKGPTSTTSKAPAPTTPKETAPTTPKEPAPTTTPK 706
Db 781 TAPTTLKEPAPTTPKKAPKELAPTTTKGPTSTTSKAPAPTTPKETAPTTPKEPAPTTTPK 840
QY 707 KPAPTTPETPTTSEVSTPTTKEPTTHKSDSTPELSABPTPKALENSPKEPGVPT 766
Db 841 KPAPTTPETPTTSEVSTPTTKEPTTHKSDSTPELSABPTPKALENSPKEPGVPT 900
QY 767 TKTPAATKPEMTTAKDKTTERDLRTTPTTTAAPKMTKETATTTKTTESKITATTTOV 826
Db 901 TKTPAATKPEMTTAKDKTTERDLRTTPTTTAAPKMTKETATTTKTTESKITATTTOV 960
QY 827 TSTTTQDTPPKITLTKTTLAPKVTTKKITTTTEIMNKPEETAKPKDRATNSKATTPK 886
Db 961 TSTTTQDTPPKITLTKTTLAPKVTTKKITTTTEIMNKPEETAKPKDRATNSKATTPK 1020
QY 887 POKETKAPKPTSTKPKMTPRVKPKTTTPKMTSTMPNLNPTSRIAEAMLQTTTRN 946
Db 1021 POKETKAPKPTSTKPKMTPRVKPKTTTPKMTSTMPNLNPTSRIAEAMLQTTTRN 1080
QY 947 QTPNSKLVEVNPKSSEDAGGAGETPHMLLRPHVMEVPTPMDYLPVNPQGIINPMLS 1006
Db 1081 QTPNSKLVEVNPKSSEDAGGAGETPHMLLRPHVMEVPTPMDYLPVNPQGIINPMLS 1140
```

```
RESULT 5
AAR26049
ID AAR26049 standard; protein; 1404 AA.
XX AC AAR26049;
XX DT 25-MAR-2003 (revised)
XX DT 02-FEB-1993 (first entry)
XX DE MSF precursor.
XX KW Megakaryocyte colony stimulating factor; secretion signal; meg-CSF;
XX KW stability; proteolytic cleavage; adhesion; alternative splicing.
XX OS Synthetic.
XX Key Location/Qualifiers
FH Region 1..26 /label= Exon_I
FT Region 26..67 /label= Exon_II
FT Region 67..107 /label= Exon_III
FT Region 107..157 /label= Exon_IV
FT Region 157..200 /label= Exon_V
FT Region 200..1141 /label= Exon_VI
FT Region 1166..1212 /label= Exon_VIII
FT Region 1213..1266 /label= Exon_IX
FT Region 1266..1331 /label= Exon_X
FT Region 1331..1373 /label= Exon_XI
FT Region 1373..1404 /label= Exon_XII
FT Region 1411..1166 /label= Exon_VII
XX WC9213075-AL.
XX 06-AUG-1992.
XX 17-JAN-1992; 92WO-US000433.
XX 18-JAN-1991; 91US-00643502.
XX 10-SEP-1991; 91US-00757022.
XX (GEMY ) GENETICS INST INC.
XX Turner K, Clark SC, Jacobs K, Hewick RM, Gesner TG;
XX WPI; 1992-284660/34.
XX N-PSDB; AAQ27223.
XX New human mega-karyocyte stimulating factors - for treating immune
XX deficiencies, cancer, exposure to radiation or drugs, bacterial and viral
XX infections, etc.
XX Claim 1, 2 and 3; Fig 1; 87pp; English.
XX The sequence given is a full length translation from the megakaryocyte
XX stimulating factor (MSF) precursor. The sequence covered by exons II, III
XX and IV encodes megakaryocyte stimulating factor (MSF). This sequence is
XX modified by the addition of an N-terminal sequence encoding a secretory
XX leader, an initiating methionine preceding exon II and a terminating
XX codon following exon IV. The cDNA sequence given contains sequences
XX derived from human megakaryocyte colony stimulating factor (meg-CSF).
XX Exon I contains the initiating methionine, and encodes a classical
XX mammalian protein secretion signal sequence. The sequence encoding the
```

original meg-CSF includes exons II-IV and is thought to terminate in the region between amino acid residues 134 - 147. The primary transcript of this gene may be cleaved in different ways to yield a family of mRNA's each encoding a different MSF protein. Exons V and VI are thought to be related to the activity of the factor and are also implicated in the stability, folding and processing of the molecule. These exons are also thought to play a role in the observed synergy of MSF with other cytokines. Exons V - XII are believed to be implicated in the processing or folding of the appropriate structure of the resulting factor, i.e. one or more of these exons may contain sequences which direct proteolytic cleavage, adhesion, organisation of the cellular matrix or extracellular matrix processing. Both naturally occurring and non-naturally occurring MSF's may be characterised by various combinations of alternatively spliced exons from this sequence, with the exons spliced together in differing orders to form different members of the MSF family. (Updated on 25-MAR-2003 to correct PN field.)

Sequence 1404 AA;

Query Match	99.4%;	Score 5339.6;	DB 2;	Length 1404;
Best Local Similarity	88.2%;	Pred. No. 2.3e-141;		
Matches 1006:	Conservative	0;	Mismatches 0;	Indels 134;
	Gaps	2;		

Qy	1	MAWKTLPIYLLLLLSVFVIQOVSSO-----	25
Db	1	MAWKTLPIYLLLLLSVFVIQOVSSQDLSSCAGRCOEGYSDATCNCDCYNQHYMECCPDF	60
Qy	26	-----ELSCKGRCFESPERGECDDAQCKYDKCCPDYESFCAE-----	65
Db	61	KRVCTAELSCKGRCFESPERGECDDAQCKYDKCCPDYESFCAEVHNPTSPSSKAP	120
Qy	66	-----	65
Db	121	PPSGASQTIKSTTKRSPKPPNNKKTKVIESBEITEHSVSENQESSSSSSSSSSSIW	180
Qy	66	-----VKONKNRPKKKPTPKPPVVDAGSLDNGDKVITPDNST	106
Db	181	KIKSGNSAANRELOKULKVXDNKKNRKKKPTPKPPVVDAGSLDNGDKVITPDNST	240
Qy	107	TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTVEKTTTTNKQTSDG	166
Db	241	TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTVEKTTTTNKQTSDG	300
Qy	167	KEKTTSAKETQSI EKTSAKOLAPTQKVIAKPTPKAETTTKGPALTTPKEPPTTPKEPAS	226
Db	301	KEKTTSAKETQSI EKTSAKDLAPTQKVIAKPTPKAETTTKGPALTTPKEPPTTPKEPAS	360
Qy	227	TTPKEPTPTTIKSAPTTPKEPATTTTKSAPTTKBPAPTTTKBPAPTTTKBP	286
Db	361	TTPKEPTPTTIKSAPTTPKEPATTTTKSAPTTKBPAPTTTKBPAPTTTKBP	420
Qy	287	APTITTKSAPTTTPKEPATTTPKKPAPTTTPKEPATTTPKGPTPTTPKEPATTTTPK	346
Db	421	APTITTKSAPTTTPKEPATTTPKKPAPTTTPKEPATTTPKGPTPTTPKEPATTTTPK	480
Qy	347	EPAPTAPKKPADTTTPKEPATTTTPKEPATTTTKBPSTTPKEPATTTKSAPTTTKEP	406
Db	481	EPAPTAPKKPADTTTPKEPATTTTKGAPTTTKBPSTTPKEPATTTKSAPTTTKEP	540
Qy	407	TTKSAPTTTPKBPSTTTTPKEPATTTTPKEPATTTPKKPAPTTTPKEPATTTTKBP	466
Db	541	TTKSAPTTTPKBPSTTTTPKEPATTTTPKEPATTTPKKPAPTTTPKEPATTTTKBP	600
Qy	467	APTAPKEPATTTPKETAPTTPKKLTPTTPEKLIAPTTPKEPADTTPEELAPTTPEEPTPT	526
Db	601	APTAPKEPATTTPKETAPTTPKKLTPTTPEKLIAPTTPKEPADTTPEELAPTTPEEPTPT	660
Qy	527	PEEPATTTPKAAAPNTPKBPAPTTTPKEPATTTTPKEPATTTPKETAPTTPKGTAPTTLKBP	586
Db	661	PEEPATTTPKAAAPNTPKBPAPTTTPKEPADTTTPKEPATTTPKETAPTTPKGTAPTTLKBP	720
Qy	587	APTTPKKPAPKELAPTTTKETPTSTTSKDPAPTTPKGTAPTTTPKEPATTTPKBPAPTTPKG	646

measuring the amount of MSF or its fragment in a biological sample of a mammal, wherein an increased amount of MSF compared to a control indicates the presence of or predisposition to developing osteoarthritis. The tribonectin and DNA encoding it are useful in the treatment of osteoarthritis, where they may be used for lubricating mammalian joints, such as articulating joints of humans, dogs or horses. The tribonectin, when formulated as a membrane, foam, gel or fibre, is useful for inhibiting adhesion between two surfaces such as the injured tissues of a mammal, where the injury is caused by a surgical insertion or trauma, or an artificial device e.g., an orthopaedic implant. In particular, one of the surfaces is pericardial tissue. DNA encoding a tribonectin may be used in gene therapy. The present sequence represents human MSF

Seq Sequence 1404 AA;

Query Match 99.4%; Score 5339.6; DB 4; Length 1404;  
 Best Local Similarity 88.2%; Pred. No. 2.3e-141;  
 Matches 1006; Conservative 0; Mismatches 0; Indels 134; Gaps 2;

QY 1 MAWKTLPIYLLLSLVFVIQVSSQ----- 25  
 DB 1 MAWKTLPIYLLLSLVFVIQVSSQDLSSCAGRCGEGSRDATCNCDYNCQHYMECCPDF 60

QY 26 -----ELSCGRCFSPRGRECDCAOQCKYDKCCPDYESFCAE----- 65  
 DB 61 KRVCTAELSCGRCFSPRGRECDCAOQCKYDKCCPDYESFCAEVHNPSTSPSSKKAP 120

QY 66 ----- 65

DB 121 PPSGASQTIKTKRSPKPPNKKTKVIESEBITTEHSVSENQESSSSSSSSSTIW 180

QY 66 -----VKONKNRTRKKKPPVVDVDEAGSLDNGDFKVTTPDTST 106  
 DB 181 KIKSSNSAANRELQKLVKONKNRTRKKKPPVVDVDEAGSLDNGDFKVTTPDTST 240

QY 107 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSUTVNKETTVEKETTNNKQSTDG 166  
 DB 241 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSUTVNKETTVEKETTNNKQSTDG 300

QY 167 KEKTTSAKETQSIKTSADLAPTSKVLAKPTPKAETTTKGNALTPPKPTTTPKEPAS 226  
 DB 301 KEKTTSAKETQSIKTSADLAPTSKVLAKPTPKAETTTKGNALTPPKPTTTPKEPAS 360

QY 227 TTPKEPTPTTIKSAPTTPKEPATTTKSAPTTKPEPATTTPKEPATTTPKEPATTTKP 286  
 DB 361 TTPKEPTPTTIKSAPTTPKEPATTTKSAPTTKPEPATTTPKEPATTTPKEPATTTKP 420

QY 287 APTTTKSAPTTPKEPATTTPKPAPTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPK 346  
 DB 421 APTTTKSAPTTPKEPATTTPKPAPTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPK 480

QY 347 EPAPTAPKPAPTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATT 406  
 DB 481 EPAPTAPKPAPTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATT 540

QY 407 TTKSAPTTPKEPSTTTTKEPATTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTKP 466  
 DB 541 TTKSAPTTPKEPSTTTTKEPATTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTKP 600

QY 467 APTAPKEPATTTPKETAPTTPKKLTPPTDEKLAPTTPEKPAPTTPPEELAPTTPEEPTPT 526  
 DB 601 APTAPKEPATTTPKETAPTTPKKLTPPTDEKLAPTTPEKPAPTTPPEELAPTTPEEPTPT 660

QY 527 PREPAPTTPKAAAPNTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEP 586  
 DB 661 PREPAPTTPKAAAPNTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEP 720

QY 587 APTTPKAPKELAPTTTKEPTSTSDKPAPTTPKGTAPTTPKEPATTTPKEPATTTPKEPATT 646  
 DB 721 APTTPKAPKELAPTTTKEPTSTSDKPAPTTPKGTAPTTPKEPATTTPKEPATTTPKEPATT 780

QY 647 TAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTSDKPAPTTPKETAPTTPKEPATTTPK 706

DB 781 TAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTSDKPAPTTPKETAPTTPKEPATTTPK 840  
 QY 707 KPAPTTPPTPTPTTSEVSTPTTKEPTTIHKSDPESTPELSABPTPKALENSPKPEGVPT 766  
 DB 841 KPAPTTPPTPTPTTSEVSTPTTKEPTTIHKSDPESTPELSABPTPKALENSPKPEGVPT 900  
 QY 767 TKTPAATKPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETATTTESKITATTATTOV 826  
 DB 901 TKTPAATKPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETATTTESKITATTATTOV 960  
 QY 827 TSTTTQDTTTPFKITTLTKTTTLAPKVTITTKKTIITTEIMNKPEETAKEPKORATNSKATTPK 886  
 DB 961 TSTTTQDTTTPFKITTLTKTTTLAPKVTITTKKTIITTEIMNKPEETAKEPKORATNSKATTPK 1020  
 QY 887 PQKPTKAPKPTSTKPKTMVRVRKPKTTTPPKMTSTMPELNPTSRIAEAMLOTTTRPN 946  
 DB 1021 PQKPTKAPKPTSTKPKTMVRVRKPKTTTPPKMTSTMPELNPTSRIAEAMLOTTTRPN 1080  
 QY 947 QTPNSKLVENVNPKSBDAGGAGETPHMLLRPHVFMPEVTPDMDYLRPVNQGIIINPMLS 1006  
 DB 1081 QTPNSKLVENVNPKSBDAGGAGETPHMLLRPHVFMPEVTPDMDYLRPVNQGIIINPMLS 1140

RESULT 7  
 AAB60568  
 ID AAB60568 standard; protein; 1404 AA.  
 XX AC AAB60568;  
 XX DT 27-APR-2001 (first entry)  
 XX DE Human megakaryocyte stimulating factor (MSF, CACP).  
 XX KW Human; CACP protein; camptodactyly-arthropathy-coxa vara-pericarditis;  
 XX KW MSF; megakaryocyte stimulating factor; synovial lubricant;  
 XX KW chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;  
 XX KW antiarthritic.  
 XX OS Homo sapiens.  
 XX XX WO200107068-A1.  
 XX PN 01-FEB-2001.  
 XX PD 21-JUL-2000; 2000WO-US020002.  
 XX PF 23-JUL-1999; 99US-0145328P.  
 XX PR 19-JUL-2000; 2000US-00145328.  
 XX XX (UYCA-) UNIV CASE WESTERN RESERVE.  
 XX PA Warman ML;  
 XX PI WPI; 2001-182721/18.  
 XX DR New composition comprising the camptodactyly-arthropathy-coxa vara-pericarditis protein in combination with an anesthetic, useful for treating osteoarthritis, or as lubricants of tissue and joints.  
 XX PS Example 1; Page; 34pp; English.  
 XX CC The invention relates to a method of treating osteoarthritis via the administration of a composition comprising the camptodactyly-arthropathy-coxa vara-pericarditis (CACP) protein, or portions of the CACP protein. The composition may further comprise a local anesthetic. The composition of the invention may be administered via intra-articular or intravenous injection. The human CACP protein is identified in the invention as being megakaryocyte stimulating factor (MSF). The gene encoding CACP protein (MSF) is located on chromosome 1q25-31, and mutations in this gene are responsible for the heritable disorder camptodactyly-arthropathy-coxa vara-pericarditis, in which patients have synovial hyperplasia without evidence of inflammation. CACP protein (MSF) acts as a synovium lubricant, and can be used to lubricate tissue and joints in the

CC treatment of osteoarthritis. The composition may be applied to reduce the  
 CC symptoms of osteoarthritis (e.g., joint pain, loss of range of movement  
 CC or joint damage). The present sequence represents human megakaryocyte  
 CC stimulating factor (MSF, CAC protein). Note: This sequence is not given  
 CC in its entirety in figure 4 of the specification, although a GenBank  
 CC accession number was given. This sequence was therefore obtained from  
 CC GenBank (U70316)

XX SQ Sequence 1404 AA;

Query Match 99.4%; Score 5339.6; DB 4; Length 1404;  
 Best Local Similarity 88.2%; Pred. No. 2.3e-141;  
 Matches 1006; Conservative 0; Mismatches 0; Indels 134; Gaps 2;

QY 1 MAWKTLPIYLLLLSVFVIOQVSSQ----- 25  
 DB 1 MAWKTLPIYLLLLSVFVIOQVSSQ----- 60

QY 26 -----ELCKGRCFESFGRGRCDDAOCKKYDKCCPDYESFCAB----- 65  
 DB 61 KRVTAEALCKGRCFESFGRGRCDDAOCKKYDKCCPDYESFCABVHNPTSPSPSKAP 120

QY 66 ----- 65  
 DB 121 PPSGASQTIKSTTRSPKPNKKTKVIBSEBITEHSVSENQESSSSSSSSSSSTIW 180

QY 66 -----VKONKNRTKKKTPKPPVDEAGSLDNGDFKTTTPTST 106  
 DB 181 KIKSSKNSAANRELQKLVKDNKNQRTKKKTPKPPVDEAGSLDNGDFKTTTPTST 240

QY 107 TOHNKYSTSPKITTAKPINRPSLPNSDTSKETSIVNKETVETKTTNNKQSTDG 166  
 DB 241 TOHNKYSTSPKITTAKPINRPSLPNSDTSKETSIVNKETVETKTTNNKQSTDG 300

QY 167 KEKTSKAKETOSIEKTSADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTTPKEPAS 226  
 DB 301 KEKTSKAKETOSIEKTSADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTTPKEPAS 360

QY 227 TTPKEPTPTTIKGAPTTPKEPAPTTTKSAPTTPKAPAPTTTPKEPAPTTTPKEPAPTTTKEP 286  
 DB 361 TTPKEPTPTTIKGAPTTPKEPAPTTTKSAPTTPKAPAPTTTPKEPAPTTTPKEPAPTTTKEP 420

QY 287 APPTTKSAPTTTPKEPAPTTTKKAPPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 346  
 DB 421 APPTTKSAPTTTPKEPAPTTTKKAPPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 480

QY 347 EPAPTAPKPEPAPTTTPKEPAPTTTPKEPAPTTTKESPPTTPKEPAPTTTKSAPTTTKEPAPT 406  
 DB 481 EPAPTAPKPEPAPTTTPKEPAPTTTPKEPAPTTTKESPPTTPKEPAPTTTKSAPTTTKEPAPT 540

QY 407 TTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKAPAPTTTPKEPAPTTTPKEPAPTTTKEP 466  
 DB 541 TTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKAPAPTTTPKEPAPTTTPKEPAPTTTKEP 600

QY 467 APAPTAPKEPAPTTTPKEPAPTTTPKLTPTTPPEKLAPTTPEKAPPTPEELAPTTPEEPPTT 526  
 DB 601 APAPTAPKEPAPTTTPKEPAPTTTPKLTPTTPPEKLAPTTPEKAPPTPEELAPTTPEEPPTT 660

QY 527 PEEPAPTTPKAAAPNTPKAPAPTTTPKEPAPTTTPKEPAPTTTPKETAPTTPKGTAPTTLKEP 586  
 DB 661 PEEPAPTTPKAAAPNTPKAPAPTTTPKEPAPTTTPKEPAPTTTPKETAPTTPKGTAPTTLKEP 720

QY 587 APPTPKAPKELAPTTTKETSTSDKAPPTTPKGTAPTTTPKEPAPTTTPKEPAPTTTPK 646  
 DB 721 APPTPKAPKELAPTTTKETSTSDKAPPTTPKGTAPTTTPKEPAPTTTPKEPAPTTTPK 780

QY 647 TAPTTLKEPAPTTTPKAPKELAPTTTKGTSTSDKAPPTTPKETAPTTTPKEPAPTTTPK 706  
 DB 781 TAPTTLKEPAPTTTPKAPKELAPTTTKGTSTSDKAPPTTPKETAPTTTPKEPAPTTTPK 840

QY 707 KPAPTTTETPTPTTSEVSTPTTTKEPTTIHKSPDESTPELSAETTPKALENSPKPEGVT 766  
 DB 841 KPAPTTTETPTPTTSEVSTPTTTKEPTTIHKSPDESTPELSAETTPKALENSPKPEGVT 900

QY 767 TKTBAATKEEMTTAKOKTTERDLRTTPTTTTAAAPKMTKETATTTTEKTTESKITATTTOV 826  
 DB 901 TKTBAATKEEMTTAKOKTTERDLRTTPTTTTAAAPKMTKETATTTTEKTTESKITATTTOV 960

QY 827 TSTTTQDTPFPFKITTLTKTTLAPKVVTTTKKTTTTEIMNKPEETAKPKORATNSKATTPK 886  
 DB 961 TSTTTQDTPFPFKITTLTKTTLAPKVVTTTKKTTTTEIMNKPEETAKPKORATNSKATTPK 1020

QY 887 POKPTKAPKPTSTKPKTTPRVKPTTTPRKMSTSTMPELNPTSRIAEAMLQTTTRN 946  
 DB 1021 POKPTKAPKPTSTKPKTTPRVKPTTTPRKMSTSTMPELNPTSRIAEAMLQTTTRN 1080

QY 947 QTPNSKLVVNPKSESDAGAGETPHMLLRPHVMEVTPDMDYLPRVNOGIIINPMLS 1006  
 DB 1081 QTPNSKLVVNPKSESDAGAGETPHMLLRPHVMEVTPDMDYLPRVNOGIIINPMLS 1140

RESULT 8  
 ADM98014  
 ID ADM98014 standard; protein; 1404 AA.  
 XX AC ADM98014;  
 XX DT 01-JUL-2004 (first entry)  
 XX DE Human megakaryocyte stimulating factor (MSF).  
 XX KW lubricating polypeptide; O-linked oligosaccharide; joint lubrication;  
 XX KW CAP; camptodactyl-arthropathy pericarditis; osteoarthritis; human;  
 XX KW megakaryocyte stimulating factor; MSF.  
 XX OS Homo sapiens.  
 XX PN US2004072741-A1.  
 XX PD 15-APR-2004.  
 XX PF 02-JUL-2001; 2001US-00897188.  
 XX PR 23-APR-1999; 99US-00298970.  
 XX PR 24-APR-2000; 2000US-00556246.  
 XX PA (JAYG/) JAY G D.  
 XX PI Jay GD;  
 XX WPi; 2004-373948/35.  
 XX DR N-PSDB; ADM98015.  
 XX PT New tribonectin polypeptides and polynucleotides for lubricating joints  
 PT or other tissues to prevent or treat camptodactyl-arthropathy-  
 PT pericarditis syndrome or osteoarthritis.  
 XX Claim 1; SEQ ID NO 1; 34pp; English.  
 XX The invention relates to a lubricating polypeptide and at least one O-  
 linked oligosaccharide. The composition and methods are useful for  
 CC lubricating joints or other tissues to prevent or treat camptodactyl-  
 CC arthropathy pericarditis (CAP) or osteoarthritis in mammals. The present  
 CC sequence represents the amino acid sequence of the human megakaryocyte  
 CC stimulating factor (MSF).  
 XX SQ Sequence 1404 AA;

Query Match 99.4%; Score 5339.6; DB 8; Length 1404;  
 Best Local Similarity 88.2%; Pred. No. 2.3e-141;  
 Matches 1006; Conservative 0; Mismatches 0; Indels 134; Gaps 2;

QY 1 MAWKTLPIYLLLLSVFVIOQVSSQ----- 25  
 DB 1 MAWKTLPIYLLLLSVFVIOQVSSQ----- 60

QY 26 -----ELCKGRCFESPERGECDDAQCKYKCCPDYESFCAE----- 65  
 Db 61 KRVTAEISCKGRCFESPERGECDDAQCKYKCCPDYESFCAEHNTPSPSSKKAP 120  
 QY 66 ----- 65  
 Db 121 PPSGASQTIKSTTKRSPKPNKKTKKVBIESEEITEHSVSENQESSSSSSSSSTIW 180  
 QY 66 -----VKDNKNRTKKKPTKPPVVDVDEAGSLONGDFKVTTPDTST 106  
 Db 181 KIKSSKNSAANRELQKKLVKDNKNRTKKKPTKPPVVDVDEAGSLONGDFKVTTPDTST 240  
 QY 107 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTIVETKETTITNKTSTDG 166  
 Db 241 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTIVETKETTITNKTSTDG 300  
 QY 167 KEKTSNAKETOSIEKTSNAKDLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTPTTPKEPAS 226  
 Db 301 KEKTSNAKETOSIEKTSNAKDLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTPTTPKEPAS 360  
 QY 227 TTPKEPTPTTIKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTKPAPTTTPKEPAPTTTKP 286  
 Db 361 TTPKEPTPTTIKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTKPAPTTTPKEPAPTTTKP 420  
 QY 287 APTTTKSAPTTTPKEPAPTTPKAPTTTPKEPAPTTTPKEPTPTTPKEPAPTTTPKEPAPTTPK 346  
 Db 421 APTTTKSAPTTTPKEPAPTTPKAPTTTPKEPAPTTTPKEPTPTTPKEPAPTTTPKEPAPTTPK 480  
 QY 347 EPAPTAPKAPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 406  
 Db 481 EPAPTAPKAPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 540  
 QY 407 TTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKP 466  
 Db 541 TTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKP 600  
 QY 467 APTAPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 526  
 Db 601 APTAPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 660  
 QY 527 REEPAPTTPKAAPNTPKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 586  
 Db 661 REEPAPTTPKAAPNTPKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 720  
 QY 587 APTTPKKPAPKELAPTTTKPSTSDKAPTTTPKGTATTPKEPAPTTTPKEPAPTTTPK 646  
 Db 721 APTTPKKPAPKELAPTTTKPSTSDKAPTTTPKGTATTPKEPAPTTTPKEPAPTTTPK 780  
 QY 647 TAPTTLKEPAPTTPKKAPKELAPTTTKGPTSTSDKAPTTTPKGTATTPKEPAPTTTPK 706  
 Db 781 TAPTTLKEPAPTTPKKAPKELAPTTTKGPTSTSDKAPTTTPKGTATTPKEPAPTTTPK 840  
 QY 707 KPAPTTPEPTPTTSEVSPPTTKPTTHKSPDSTPELSAEPPTPKALENPKGPGVPT 766  
 Db 841 KPAPTTPEPTPTTSEVSPPTTKPTTHKSPDSTPELSAEPPTPKALENPKGPGVPT 900  
 QY 767 TKTPAANKPEMTTAKDKTTERDLRTPTTTTAAAPKMTKETATTTKTTESKITATTQV 826  
 Db 901 TKTPAANKPEMTTAKDKTTERDLRTPTTTTAAAPKMTKETATTTKTTESKITATTQV 960  
 QY 827 TSTTTQDTPPKITLTKTTTLAPKVTTKKTTITTEIMNKPEETAKPKDRATNSKATTPK 886  
 Db 961 TSTTTQDTPPKITLTKTTTLAPKVTTKKTTITTEIMNKPEETAKPKDRATNSKATTPK 1020  
 QY 887 POKPTKAPKPTSTTKPKTMPVRPKTTPTRKMTSNMPELNPTSRIAEAMLQTTTPN 946  
 Db 1021 POKPTKAPKPTSTTKPKTMPVRPKTTPTRKMTSNMPELNPTSRIAEAMLQTTTPN 1080  
 QY 947 QTPNSKLVENPKSDAGGAEGETHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS 1006  
 Db 1081 QTPNSKLVENPKSDAGGAEGETHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS 1140

RESULT 9  
 AAU32262  
 ID AAU32262 standard; protein; 1415 AA.  
 XX AC AAU32262;  
 XX AC  
 DT 18-DEC-2001 (first entry)  
 XX  
 DE Novel human secreted protein #2753.  
 XX  
 KW Human; vaccination; gene therapy; nutritional supplement;  
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;  
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.  
 OS Homo sapiens.  
 XX  
 FN WC200179449-A2.  
 XX  
 PD 25-OCT-2001.  
 XX  
 PF 16-APR-2001; 2001WO-US0008656.  
 XX  
 PR 18-APR-2000; 2000US-00552929.  
 PR 26-JAN-2001; 2001US-00770160.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Drmanac RT;  
 XX  
 DR WPI; 2001-611725/70.  
 XX  
 PT Nucleic acids encoding a range of human polypeptides, useful in genetic  
 PT vaccination, testing and therapy.  
 XX  
 PS Claim 20; Page 573; 765pp; English.  
 XX  
 CC The invention relates to novel human secreted polypeptides. The  
 CC polypeptides and antibodies to the polypeptides are useful for  
 CC determining the presence of or predisposition to a disease associated  
 CC with altered levels of polypeptide. The polypeptides are also useful for  
 CC identifying agents (agonists and antagonists) that bind to them. Cells  
 CC expressing the proteins are useful for identifying a therapeutic agent  
 CC for use in treatment of a pathology related to aberrant expression or  
 CC physiological interactions of the polypeptide. Vectors comprising the  
 CC nucleic acids encoding the polypeptides and cells genetically engineered  
 CC to express them are also useful for producing the proteins. The proteins  
 CC are useful in genetic vaccination, testing and therapy, and can be used  
 CC as nutritional supplements. They may be used to increase stem cell  
 CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon  
 CC and/or nerve tissue growth or regeneration; immune suppression and/or  
 CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.  
 CC AAU29510-AAU3304 represent the amino acid sequences of novel human  
 CC secreted proteins of the invention  
 XX  
 SQ Sequence 1415 AA;

Query Match 99.4%; Score 5339.6; DB 4; Length 1415;  
 Best Local Similarity 88.2%; Pred. No. 2.3e-141;  
 Matches 1006; Conservative 0; Mismatches 0; Indels 134; Gaps 2;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25  
 Db 12 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSDRATNCNDYNCOHYMBCCPDF 71  
 QY 26 -----ELCKGRCFESPERGECDDAQCKYKCCPDYESFCAE----- 65  
 Db 72 KRVTAEISCKGRCFESPERGECDDAQCKYKCCPDYESFCAEHNTPSPSSKKAP 131  
 QY 66 ----- 65  
 Db 132 PPSGASQTIKSTTKRSPKPNKKTKKVBIESEEITEHSVSENQESSSSSSSSSTIW 191  
 QY 66 -----VKDNKNRTKKKPTKPPVVDVDEAGSLONGDFKVTTPDTST 106



Db 192 KIKSSKNSAANRELQKKLVKDNKNRKKTKPPVVDVDEAGSLDNGDFKVTTPDNT 251  
Qy 107 TOHNVSTSPKLTITAKPINRSLPNSDTSKETSIVNKEITVETKETTINKQSTDG 166  
Db 252 TOHNVSTSPKLTITAKPINRSLPNSDTSKETSIVNKEITVETKETTINKQSTDG 311  
Qy 167 KEKTTSAKETQSIKTSKDLAPTSKVLAKPTPKAETTTKGPALTTPKSPPTTPKEPAS 226  
Db 312 KEKTTSAKETQSIKTSKDLAPTSKVLAKPTPKAETTTKGPALTTPKSPPTTPKEPAS 371  
Qy 227 TTPKEPTPTTIKSAPTTKPEPAPTTTKSAPTTKPEPAPTTTKPEPAPTTTKPE 286  
Db 372 TTPKEPTPTTIKSAPTTKPEPAPTTTKSAPTTKPEPAPTTTKPEPAPTTTKPE 431  
Qy 287 APTTTKSAPTTKPEPAPTTPKKAPTTKPEPAPTTKPEPTPTPKPEPAPTTKPE 346  
Db 432 APTTTKSAPTTKPEPAPTTPKKAPTTKPEPAPTTKPEPTPTPKPEPAPTTKPE 491  
Qy 347 EPAPTAPKPKAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKSAPTTKPE 406  
Db 492 EPAPTAPKPKAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTTKSAPTTKE 551  
Qy 407 TTKSAPTTKPEPAPTTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTTKKP 466  
Db 552 TTKSAPTTKPEPAPTTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTTKKP 611  
Qy 467 APTAPKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTPEELAPTTPEE 526  
Db 612 APTAPKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTPEELAPTTPEE 671  
Qy 527 PESAPTTKAAAPNTPKPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKGTAPTTKPE 586  
Db 672 PESAPTTKAAAPNTPKPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKGTAPTTKPE 731  
Qy 587 APTTPKPKAPKELAPTTKPEPAPTTSDKAPTTKGTAPTTKPEPAPTTKPEPAPTTK 646  
Db 732 APTTPKPKAPKELAPTTKPEPAPTTSDKAPTTKGTAPTTKPEPAPTTKPEPAPTTK 791  
Qy 647 TAPTTLKEPAPTTKPKAPKELAPTTTKGPTSTTSKAPTTPKETAPTTKPEPAPTTK 706  
Db 792 TAPTTLKEPAPTTKPKAPKELAPTTTKGPTSTTSKAPTTPKETAPTTKPEPAPTTK 851  
Qy 707 KPAPTTPEPPTTSEVSTPTTKPEPTTIHKSDESTPELSAPTPKALENSPKEPVPT 766  
Db 852 KPAPTTPEPPTTSEVSTPTTKPEPTTIHKSDESTPELSAPTPKALENSPKEPVPT 911  
Qy 767 TKTPAATKPEMTTAKDKTTERDLRTTPTTAAAPKMTKETATTTKTESKITATTTQV 826  
Db 912 TKTPAATKPEMTTAKDKTTERDLRTTPTTAAAPKMTKETATTTKTESKITATTTQV 971  
Qy 827 TSITTTQDTTTPFKITLTKTTLAPKVTITTKITITTEIMNKPEETAAPKDRATNSKATTPK 886  
Db 972 TSITTTQDTTTPFKITLTKTTLAPKVTITTKITITTEIMNKPEETAAPKDRATNSKATTPK 1031  
Qy 887 PQKETAAPKPTSTKPKTPRVRUKPTTTPRKTSTMPKELAPTSIAEAMLOTITPRN 946  
Db 1032 PQKETAAPKPTSTKPKTPRVRUKPTTTPRKTSTMPKELAPTSIAEAMLOTITPRN 1091  
Qy 947 QTNKSLVENVNPKSEDAGGAEGETPHMLLRPHVEMPEVTPDMDYLRVPVNOGIIINPMLS 1006  
Db 1092 QTNKSLVENVNPKSEDAGGAEGETPHMLLRPHVEMPEVTPDMDYLRVPVNOGIIINPMLS 1151

RESULT 10

ID ADK65839

XX ADK65839 standard; protein; 1404 AA.

AC ADK65839;

XX AC

XX 06-MAY-2004 (first entry)

XX ADK65839; Angiogenesis-differentially expressed protein #53.

XX cytostatic; cardiant; vasotropic; antiarteriosclerotic;  
KW angiogenesis inhibitor; angiogenesis stimulator; angiogenic index;  
KW gene expression; cancer; coronary artery disease; myocardial ischemia;  
KW coronary arteriosclerosis; forensic medicine.  
OS Homo sapiens.  
XX WO2003066831-A2.  
XX 14-AUG-2003.  
XX 07-FEB-2003; 2003WO-US003848.  
XX 07-FEB-2002; 2002US-00067482.  
XX 10-JUN-2002; 2002US-00164595.  
XX 16-AUG-2002; 2002US-0403649P.  
XX 03-JAN-2003; 2003US-0437746P.  
XX (ORIG-) ORIGENE TECHNOLOGIES INC.  
XX Sun Z, Li X, Kovacs KF, Fan W, Jay G;  
XX WPI; 2003-731502/69.  
XX Determining the angiogenic index of a tissue or cell sample using  
XX expression levels of differentially expressed genes, useful for  
XX diagnosing or treating cancer, coronary artery disease, myocardial  
XX ischemia and/or arteriosclerosis.  
XX Disclosure; SEQ ID NO 78; 296pp; English.  
XX The invention relates to a method of determining the angiogenic index of  
XX a tissue or cell sample comprising assessing, in a sample, the expression  
XX levels of one or more differentially-expressed gene from any of 34 DNA  
XX sequences, given in the specification, where the levels are indicative of  
XX the angiogenic index. The methods and compositions of the present  
XX invention are useful for diagnosing, preventing and/or treating cancer,  
XX coronary artery disease, myocardial ischemia or coronary  
XX arteriosclerosis. They can also be used in research, drug discovery and  
XX forensic medicine involving angiogenesis. This sequence corresponds to  
XX one of the differentially expressed proteins of the invention.  
XX  
XX Sequence 1404 AA;  
Query Match 99.2%; Score 5330.6; DB 7; Length 1404;  
Best Local Similarity 88.1%; Pred. No. 4.1e-141;  
Matches 1004; Conservative 0; Mismatches 2; Indels 134; Gaps 2;  
Qy 1 MAMKTLPIYLLLLSVFVIOQVSSQ----- 25  
Db 1 MAMKTLPIYLLLLSVFVIOQVSSQDLSSCAGRGEGYSRATCNCNDYNCQHYMECCPDF 60  
Qy 26 -----ELSCKGRCFESFERGRCDDAQCKYDKCCPDYESPCAE----- 65  
Db 61 KRVCTAELSCKGRCFESFERGRCDDAQCKYDKCCPDYESFCAEVHNFTSPSSKKAP 120  
Qy 66 ----- 65  
Db 121 PPSGASOTIKSTTKRSPKPNKKTKVIESEITEHVSSENQSSSSSSSSSSSTIR 180  
Qy 66 -----VKDNKNRRTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDNT 106  
Db 181 KIKSSKNSAANRELQKKLVKDNKNRRTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDNT 240  
Qy 107 TOHNVSTSPKLTITAKPINRSLPNSDTSKETSIVNKEITVETKETTINKQSTDG 166  
Db 241 TOHNVSTSPKLTITAKPINRSLPNSDTSKETSIVNKEITVETKETTINKQSTDG 300  
Qy 167 KEKTTSAKETQSIKTSKDLAPTSKVLAKPTPKAETTTKGPALTTPKSPPTTPKEPAS 226  
Db 301 KEKTTSAKETQSIKTSKDLAPTSKVLAKPTPKAETTTKGPALTTPKSPPTTPKEPAS 360





```

ABUS3253
ID ABUS3253 standard; protein; 551 AA.
XX
AC ABUS3253;
XX
DT 14-APR-2003 (first entry)
XX
DE Human testes-derived DKFphtes3_4019 homologue #2.
XX
KW Human; gene therapy; vaccine; disease treatment; detection.
XX
OS Homo sapiens.
XX
PN WO200112659-A2.
XX
PD 22-FEB-2001.
XX
XX 18-AUG-2000; 2000WO-IB001496.
XX
PR 18-AUG-1999; 99US-0149499P.
PR 28-SEP-1999; 99US-0156503P.
XX
PA (GEHU-) GERMAN HUMAN GENOME PROJECT.
XX
PI Wiemann S;
XX
DR WPI; 2001-327840/34.
XX
PT Nucleic acids having the sequences of clones isolated from libraries of
PT different human tissues, useful in recombinant DNA methodologies.
XX
PS Example III; Page 892-893; 1095pp; English.
XX
CC This invention describes novel polynucleotides and polypeptides isolated
CC from human cDNA libraries which can be used for gene therapy or in
CC vaccines. The polynucleotides of the invention and antibodies encoded by
CC them may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate polypeptide expression. The products of the
CC invention may also be used to identify modulators of expression and
CC activity and to down regulate expression and activity. The antibodies of
CC the invention may also be used as diagnostic agents for detecting the
CC presence of polypeptides in samples. This sequence represents a homologue
CC of a polypeptide described in the disclosure of the invention
XX
SQ Sequence 551 AA;

Query Match 53.8%; Score 2888; DB 4; Length 551;
Best Local Similarity 100.0%; Pred. No. 2.7e-73;
Matches 544; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 463 TKKPAPAPKEPAPTPKETAPTPKGLTPTTPEKLAPTPPEKPAPTPPELAPTPPEP 522
DB 1 TKKPAPAPKEPAPTPKETAPTPKGLTPTTPEKLAPTPPEKPAPTPPELAPTPPEP 60

QY 523 TPTTPEPAPTPPKAAAPNTPKEPAPTPPKAPAPTPPKETAPTPPKGTAPTT 582
DB 61 TPTTPEPAPTPPKAAAPNTPKEPAPTPPKAPAPTPPKETAPTPPKGTAPTT 120

QY 583 LKEPAPTPPKAPKELAPTTTKEPTSTTSKAPAPTPPKGTAPTPKEPAPTPKEPAPT 642
DB 121 LKEPAPTPPKAPKELAPTTTKEPTSTTSKAPAPTPPKGTAPTPKEPAPTPKEPAPT 180

QY 643 TPKGATPTLLKEPAPTPPKAPKELAPTTTKEPTSTTSKAPAPTPPKETAPTPKEPAP 702
DB 181 TPKGATPTLLKEPAPTPPKAPKELAPTTTKEPTSTTSKAPAPTPPKETAPTPKEPAP 240

QY 703 TTPKKAPAPTPPPPTTSVSTPTTKEPTTIHKSPDESTPBLSAEPTPKALENSPKPEP 762
DB 241 TTPKKAPAPTPPPPTTSVSTPTTKEPTTIHKSPDESTPBLSAEPTPKALENSPKPEP 300

QY 763 GVPPTTKTAAATKPEMTTAKDKITTERDLRTTPETTTAAPKMTKETATTTTCKITAT 822
DB 301 GVPPTTKTAAATKPEMTTAKDKITTERDLRTTPETTTAAPKMTKETATTTTCKITAT 360

```

```

QY 823 TTQVSTTTQDTPPKITTLKTTILAPKVTITTKITTTITIMNKPETAKPKDRATNSKA 882
DB 361 TTQVSTTTQDTPPKITTLKTTILAPKVTITTKITTTITIMNKPETAKPKDRATNSKA 420
QY 883 TTPKPKQPTKAPKKPTSTTKPKTMPRVKPKTTTPPRKMTSTMPPELNPTSRIAEAMLOTT 942
DB 421 TTPKPKQPTKAPKKPTSTTKPKTMPRVKPKTTTPPRKMTSTMPPELNPTSRIAEAMLOTT 480
QY 943 TRPNQTPNSKLVEVNPKSEDAAGAGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIIN 1002
DB 481 TRPNQTPNSKLVEVNPKSEDAAGAGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIIN 540
QY 1003 PMLS 1006
DB 541 PMLS 544

RESULT 13
AAB29778
ID AAB29778 standard; protein; 902 AA.
XX
AC AAB29778;
XX
DT 28-FEB-2001 (first entry)
XX
DE Human MSF-derived tribonectin.
XX
KW Human tribonectin; MSF; megakaryocyte stimulating factor;
KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;
KW osteoarthritis; tribosupplementation; tissue adhesion inhibition;
KW friction coefficient reduction; gene therapy; antiarthritic; osteopathic.
XX
OS Homo sapiens.
XX
PN WO200064930-A2.
XX
PD 02-NOV-2000.
XX
PF 24-APR-2000; 2000WO-US010953.
XX
PR 23-APR-1999; 99US-00298970.
XX
PA (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.
XX
PI Jay GD;
XX
DR WPI; 2001-024673/03.
XX
PT Novel tribonectin polypeptide useful as lubricant for treating
PT osteoarthritis, comprises O-linked lubricating moiety.
XX
PS Disclosure; Fig 1; 47pp; English.
XX
CC The invention relates to a human tribonectin which is a product of
CC alternative splicing of the human MSF (megakaryocyte stimulating factor)
CC gene. The tribonectin has at least one O-linked oligosaccharide
CC lubricating moiety and has a polypeptide sequence comprising 1-76 repeats
CC of a motif having at least 50% identity to the sequence KEPAPPT
CC (AAB29774). The invention also relates to a nucleic acid encoding a human
CC MSF-derived tribonectin; a biocompatible composition comprising a human
CC tribonectin for inhibiting tissue adhesion formation; and a method of
CC diagnosing osteoarthritis or a predisposition to osteoarthritis by
CC measuring the amount of MSF or its fragment in a biological sample of a
CC mammal, wherein an increased amount of MSF compared to a control
CC indicates the presence of or predisposition to developing osteoarthritis.
CC The tribonectin and DNA encoding it are useful in the treatment of
CC osteoarthritis, where they may be used for lubricating mammalian joints,
CC such as articulating joints of humans, dogs or horses. The tribonectin,
CC when formulated as a membrane, foam, gel or fibre, is useful for
CC inhibiting adhesion between two surfaces such as the injured tissues of a
CC mammal, where the injury is caused by a surgical insertion or trauma, or
CC an artificial device e.g., an orthopaedic implant. In particular, one of

```

CC the surfaces is pericardial tissue. DNA encoding a tribonectin may be  
 CC used in gene therapy. The present sequence represents a substantial  
 CC portion of a human MSP-derived tribonectin

XX SQ Sequence 902 AA;

Query Match 53.6%; Score 2880.1; DB 4; Length 902;  
 Best Local Similarity 71.4%; Pred. No. 7.9e-73;  
 Matches 675; Conservative 27; Mismatches 74; Indels 169; Gaps 66;

QY 66 VKONKKNRTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDSTTQHNVKSTSPKITTAKPIN 125  
 Db 1 VKONKKNRTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDSTTQHNVKSTSPKITTAKPIN 60

QY 126 PRSLPNSDTSKETSILTWNKETTIVETKETTNNKQSTSDGKEKTTSAKETQSIKTSK 185  
 Db 61 PRSLPNSDTSKETSILTWNKETTIVETKETTNNKQSTSDGKEKTTSAKETQSIKTSK 120

QY 186 DLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKGPPTTTIKSAPTTPK 245  
 Db 121 DLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKGPPTTTIKSAPTTPK 172

QY 246 EPAPTTTKSAPTTKGPAPTTTKAPATTTPKEPATTTPKEPATTTPKSAPTTPKEPAPT 305  
 Db 173 EPAPTTTKSAPTTKGPAP--TTKEPAPTT--KEPAP--TTKEPAPTT--KEPAPT 228

QY 306 PKKEPAPTTKEPAPTTPKPTPTTPKEPATTTPKEPATTTPKEPAPTAPKKPAPTTKEPA 365  
 Db 229 --KEPAPT--KEPAPT--KEPAPT--KEPAPTTPKEPATT--KEPAPT--KEPAPT--KEPA 281

QY 366 PTPKPEPAPTTTKPSPTTPKEPATTTPKSAPTTTKEPATTTPKSAPTTPKPSPTTKE 425  
 Db 282 PTT--KEPAP--TTKEPAPT--KEPAPTTPKEPAP--TTKEPAPTTPKEPATT--KEPAP--TTKE 335

QY 426 PAPTTPKEPAPTTKKPAPTTTPKEPATTTPKEPATTTPKAPATPKAPAPTTPKETAPT 485  
 Db 336 PAPTTP--KEPAPT--KEPAPT--KEPAPT--KEPAP--TTKEPAPT--KEPAPT--KEPAPT 388

QY 486 TPKKLTPTTPEKLAPTTPEKAPTTPEELAPTTPEPTTPPEPAPTTPKAAAPTPKE 545  
 Db 389 T--KEPAPTTPKEP--APTTPKE--PAPTTP--KEPAPT--KEPAPT--KEPAPT--KE 440

QY 546 PAPTTPKEPAPTTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKAP--KELAPT 603  
 Db 441 PAPTTP--KEPAPT--KEPAPT--KEPAPT--KEPAPT--KEPAPT--KEPAPTTPKEPAP--T 493

QY 604 TKEPTSTTSKAPATTTPKGTAPTTTPKEPAPTTTPKEPAPTTPKGTAPTTLKEPAPTTPKKP 663  
 Db 494 TKEPAPTTPKE--PAPTTP--KEPAPT--KEPAPT--KEPAPT--KEPAPT--KEPAPT--KEP 546

QY 664 AP--KELAPTTPKGTSTTSKAPATTTPKETAPTTPKEPAPTTPKKAPATTTPPTPTTS 721  
 Db 547 APTTPKEPAP--TTKEPAPTTPKE--PAPTTP--KEPAPT--KEPAPT--KEPAPTTPKE--PAPTTPK 600

QY 722 EVSTPTPTTKEPTTIHKSPOESTPELSAEPTPKALENSPKEGVPVTKTIPAAKPEMTTUA 781  
 Db 601 EPA--PTTKEPAPTTPKEPAPTTP--KEPAP--TTTKEP--APTTPKEPAPT-- 640

QY 782 KOKTBERDLRTTPETTAAAPKMTKEPATTTTEKTTESKITATTTQVSTSTTQDTPPKIIT 841  
 Db 641 -----TTKEPAPTTP--TTKEPAPTTP--TTKEPAPTTP--TTKEPAPTTP--TTKEPAPTTP 667

QY 842 LKTTTLAPKVTTTKTITTEIMNKBEETAKPKDRATNSKAITPKQKTPKAPKKTSTK 901  
 Db 668 -----TTKEPAPTTPKEPAPTTP--TTKEPAPTTPKEPAPTTP--TTKEPAPTTPKEPAPTTP 694

QY 902 KPKTMPVRKPKTTPTPRKWSTMPBELNPTSRIAEAMLQTTTPRNOTNSKLVENPKSE 961  
 Db 685 EP-----APTTPTPRKWSTMPBELNPTSRIAEAML--TTTPRNOTNSKLVENPKSE 735

QY 962 DAGGAEGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS 1006  
 Db 736 DAGGAEGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS 780

## RESULT 14

ABU53254  
 ID ABU53254 standard; protein; 513 AA.

XX AC ABU53254;  
 XX DT 14-APR-2003 (first entry)  
 XX DE Human testes-derived DKFZphtes3\_4019 homologue #3.  
 XX KW Human; gene therapy; vaccine; disease treatment; detection.

XX OS Homo sapiens.

XX PN WO200112659-A2.

XX PD 22-FEB-2001.

XX PF 18-AUG-2000; 2000WO-IB001496.

XX PR 18-AUG-1999; 99US-0149499P.

XX PR 28-SEP-1999; 99US-0156503P.

XX PA (GEHU-) GERMAN HUMAN GENOME PROJECT.

XX XX Wiemann S;

XX PI WPI; 2001-327840/34.

XX DR Nucleic acids having the sequences of clones isolated from libraries of  
 different human tissues, useful in recombinant DNA methodologies.

XX PT Example III; Page 893; 1095pp; English.

XX CC This invention describes novel polynucleotides and polypeptides isolated  
 from human cDNA libraries which can be used for gene therapy or in  
 vaccines. The polynucleotides of the invention and antibodies encoded by  
 them may be used in the prevention, diagnosis and treatment of diseases  
 associated with inappropriate polypeptide expression. The products of the  
 invention may also be used to identify modulators of expression and  
 activity and to down regulate expression and activity. The antibodies of  
 the invention may also be used as diagnostic agents for detecting the  
 presence of polypeptides in samples. This sequence represents a homologue  
 of a polypeptide described in the disclosure of the invention

XX SQ Sequence 513 AA;

Query Match 51.3%; Score 2757; DB 4; Length 513;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-69;  
 Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 RTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDSTTQHNVKSTSPKITTAKPINPRSLPP 132  
 Db 1 RTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDSTTQHNVKSTSPKITTAKPINPRSLPP 60

QY 133 NSDTSKETSILTWNKETTIVETKETTNNKQSTSDGKEKTTSAKETQSIKTSKADLAPT 192  
 Db 61 NSDTSKETSILTWNKETTIVETKETTNNKQSTSDGKEKTTSAKETQSIKTSKADLAPT 120

QY 193 VLAKPTPKAETTTKGPALTTPKEPTTPKEPASTTPKGPPTTTIKSAPTTPKEPAPT 252  
 Db 121 VLAKPTPKAETTTKGPALTTPKEPTTPKEPASTTPKGPPTTTIKSAPTTPKEPAPT 180

QY 253 KSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKSAPTTPKAPAPT 312  
 Db 181 KSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKSAPTTPKAPAPT 240

QY 313 TPKEPAPTTPKEPTTPTPKEPAPTTPKEPAPTTPKAPAPTTPKAPAPTTPKAPAPT 372  
 Db 241 TPKEPAPTTPKEPTTPTPTPKEPAPTTPKEPAPTTPKAPAPTTPKAPAPTTPKAPAPT 300

QY 373 APTTTKEPSPPTPKAPATTIKSAPTTTKEPAPTITKSAPTTTPKESPTTKEPAPTTPK 432  
 Db |||||  
 QY 301 APTTKEPSPPTPKAPATTIKSAPTTTKEPAPTITKSAPTTTPKESPTTKEPAPTTPK 360  
 Db |||||  
 QY 433 EPAPTTPKAPATTPKAPATTTPKAPATTTPKAPATTTPKAPATTTPKAPATTTPK 492  
 Db |||||  
 QY 361 EPAPTTPKAPATTPKAPATTTPKAPATTTPKAPATTTPKAPATTTPKAPATTTPK 420  
 Db |||||  
 QY 493 TTPKEKAPATTTPKAPATTTPKAPATTTPKAPATTTPKAPATTTPKAPATTTPK 552  
 Db |||||  
 QY 421 TTPKEKAPATTTPKAPATTTPKAPATTTPKAPATTTPKAPATTTPKAPATTTPK 480  
 Db |||||  
 QY 553 EPAPTTPKAPATTPKAPATTTPKAPATTTPKAPATTTPKAPATTTPKAPATTTPK 585  
 Db |||||  
 QY 481 EPAPTTPKAPATTPKAPATTTPKAPATTTPKAPATTTPKAPATTTPKAPATTTPK 513  
 Db |||||

## RESULT 15

AAR80041

ID AAR80041 standard; protein; 452 AA.

XX AAR80041;

XX 25-MAR-2003 (revised)

DT 10-APR-1996 (first entry)

XX Human megakaryocytopoietin protein.

XX Human; megakaryocytopoietin; wheat germ agglutinin; heparin;

KW megakaryocyte; aplastic anaemia; thrombocytopenia; hematocytopenia;

XW multipotential stem cell.

XX Homo sapiens.

XX Key Location/Qualifiers

FH Misc-difference 393..396

FT /note= "unspecified amino acids"

FT Misc-difference 444..446

FT /note= "unspecified amino acids"

XX W09523861-AL.

XX 08-SEP-1995.

XX 06-MAR-1995; 95WO-CN000015.

XX 04-MAR-1994; 94CN-00112066.

XX (SHAN-) SHANGHAI BEITE BIOTECHNOLOGY CO LTD.

XX Gu X, Han Z, Shen Q;

XX WPI; 1995-320576/41.

XX N-PSDB; AAT04546.

XX New haematopoietic cell growth factor - used for treating

PT thrombocytopenia and hematocytopenia.

XX Example; Page 23; 36pp; Chinese.

XX This sequence represents the human megakaryocytopoietin (MPO) protein.  
 CC This sequence was purified using a carrier which can couple wheat germ  
 CC agglutinin and heparin to separate MPO. Fragments of this sequence (see  
 CC AAR80039 and AAR80040) were used to produce the amplification primers  
 CC shown in AAT04544 and AAT04545. The fragments amplified by these primers  
 CC can then be used as probes to screen human cDNA libraries for MPO cDNA.  
 CC The MPO cDNA can then be inserted into a plasmid which is used to  
 CC transform cells to produce MPO. The MPO sequence is capable of promoting  
 CC colony formation of megakaryocytes, enlarging the size of megakaryocytes  
 CC and stimulating the proliferation of multipotential stem cells. The  
 CC factor may be used for treating thrombocytopenia and hematocytopenia. The  
 CC purification method can be used to isolate MPO from human urine or serums  
 CC of patients with aplastic anaemia, and from animal blood or urine by

CC radiation exposing the animals to induce aplastic anaemia. (Updated on 25  
 CC -MAR-2003 to correct PA field.)

SQ Sequence 452 AA;

Query Match 31.5%; Score 1694.8; DB 2; Length 452;

Best Local Similarity 72.2%; Pred. No. 4.8e-40;

Matches 328; Conservative 3; Mismatches 21; Indels 102; Gaps 3;

QY 1 MAWKTLPIYLLLLLSVFVIQQVSSQ----- 25

Db 1 MAWKTLPIYLLLLLSVFVIQQVSSQDLSSCAGRCGEGYGRDATCNCYNCQHYMECCPDF 60

QY 26 -----ELSCGRCFSPERGRECDCAOCKYDKCCPDYESFCA----- 64

Db 61 KRVCTAELSCGRCFSPERGRECDCAOCKYDKCCPDYESFCAEVHNFTSPSSKKAP 120

QY 65 -----FVKONKKNRTKKKTPKPPVVVDEAG 89

Db 121 PPSGASQTIKSTTKRSPKPPNKKTKKVIESEITEVKONKKNRTKKKTPKPPVVVDEAG 180

QY 90 SGLDNGDPKVTTPDTSSTOHNKVSTSPKITTAKPINRPSPSLPNSDTSKETSILTANKETT 149

Db 181 SGLDNGDFKVTTPDTSSTOHNKVSTSPKITTAKPINRPSPSLPNSDTSKETSILTANKETT 240

QY 150 VETKETTITNKOTSTDGKEKTTSAKETQSIKTSADKLAPTSKVLAKPTPKAETTTKGPA 209

Db 241 VETKETTITNKOTSTDGKEKTTSAKETQSIKTSADKLAPTSKVLAKPTPKAETTTKGPA 300

QY 210 LTTPKETPTTPKPEPASTTPKBPPTTIKSAPTTPKPEPAPTTTKSAPTTPKPEPAPTTTKE 269

Db 301 LTTPKETPTTPKPEPASTTPKBPPTTIKSAPTTPKPEPAPTTTKSAPTTPKPEPAPTTTKE 360

QY 270 PAPTTPKPEPAPTTTPKPEPAPTTTKSAPTTPKPEPAPTTPKKPAPTTPKPEPAPTTTKEPTT 329

Db 361 PAPTTPKPEPAPTTTPKPEPAPTTTKSHPLPRSCXXXCTQP-----TPKEHPPL 409

QY 330 FKEPAPTTTPKPEPAPTTTPKPEPAPTTAPKKPAPTTPKX 363

Db 410 PRSLHPTTPKPEPAPTTTPKPEPAPTTAPKKPAPLPPL 443

Search completed: October 13, 2004, 11:37:08

Job time : 90.1065 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 17.4873 Seconds  
(without alignments)  
3815.116 Million cell updates/sec

Title: SEQ1-E

Perfect score: 5373

Sequence: 1 MAWKTLPIYLLLSVFVIQ.....DMDYLRVFNQGIINPMLS 1006

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.1

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents: AA:\*

1: /cgn2\_6/prodata/1/iaa/5A COMB.pep:\*

2: /cgn2\_6/prodata/1/iaa/5B COMB.pep:\*

3: /cgn2\_6/prodata/1/iaa/6A COMB.pep:\*

4: /cgn2\_6/prodata/1/iaa/6B COMB.pep:\*

5: /cgn2\_6/prodata/1/iaa/PCITUS COMB.pep:\*

6: /cgn2\_6/prodata/1/iaa/backfilesi.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5373	100.0	1038	4	US-07-757-022B-74
2	5373	100.0	1270	4	US-07-757-022B-44
3	5358.9	99.7	1311	4	US-07-757-022B-42
4	5358.7	99.7	1049	4	US-07-757-022B-58
5	5358.7	99.7	1313	4	US-07-757-022B-142
6	5358	99.7	1320	4	US-07-757-022B-46
7	5358	99.7	1320	4	US-07-757-022B-60
8	5353.7	99.6	1363	4	US-07-757-022B-52
9	5349	99.6	1320	4	US-10-164-595-58
10	5344.6	99.5	1354	4	US-07-757-022B-48
11	5343.9	99.5	1361	4	US-07-757-022B-40
12	5339.6	99.4	1140	4	US-07-757-022B-104
13	5339.6	99.4	1404	4	US-07-757-022B-2
14	5339.6	99.4	1404	4	US-07-757-022B-62
15	5338.6	99.4	1404	4	US-09-298-970A-1
16	5330.6	99.2	1404	4	US-10-164-595-78
17	5241.8	97.6	1022	4	US-07-757-022B-84
18	5232.6	97.4	1314	4	US-07-757-022B-50
19	5011	93.3	941	4	US-07-757-022B-14
20	1335.3	24.9	5179	4	US-09-538-092-1258
21	1075.7	20.0	8991	4	US-08-714-741-32
22	803.4	15.0	3256	4	US-09-919-172-98
23	803.4	15.0	3256	4	US-09-976-594-22
24	803.4	15.0	3256	4	US-09-919-039-21
25	795	14.8	3118	3	US-09-578-181-1
26	790.9	14.7	2972	3	US-09-579-181-2
27	757.4	14.1	1837	3	US-08-928-361B-5

Sequence 5, Appli  
Sequence 1142, Ap  
Sequence 6, Appli  
Sequence 6, Appli  
Sequence 5, Appli  
Sequence 4, Appli  
Sequence 30227, A  
Sequence 4, Appli  
Sequence 3, Appli  
Sequence 17307, A  
Sequence 425, App  
Sequence 3, Appli  
Sequence 726, App  
Sequence 1135, Ap  
Sequence 1262, Ap  
Sequence 5, Appli  
Sequence 98, Appl  
Sequence 110, App

#### ALIGNMENTS

##### RESULT 1

US-07-757-022B-74  
; Sequence 74, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/07757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 08-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 74:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1038 amino acids  
; TYPE: AMINO ACID

TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-757-022B-74

Query Match 100.0%; Score 5373; DB 4; Length 1038;  
Best Local Similarity 100.0%; Pred. No. 1.5e-167;  
Matches 1006; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQELSCKGRCFESFERGRCDCDAQCKKYDKCCPDYE 60  
DB 1 MAWKTLPIYLLLLSVFVIQVSSQELSCKGRCFESFERGRCDCDAQCKKYDKCCPDYE 60

QY 61 SFCAEVKDNKNRTKKKPTKPPVDEAGSLDNGDFKVTTPDTSTTQHKNVSTSPKITT 120  
DB 61 SFCAEVKDNKNRTKKKPTKPPVDEAGSLDNGDFKVTTPDTSTTQHKNVSTSPKITT 120

QY 121 AKPINRPSLPNSDTSKETSITVKNKETTVEKTTTNNKQTSIDGKEKTTSAKETQSIE 180  
DB 121 AKPINRPSLPNSDTSKETSITVKNKETTVEKTTTNNKQTSIDGKEKTTSAKETQSIE 180

QY 181 KTSADLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTTPKEPASTTTPKEPTTTIKSA 240  
DB 181 KTSADLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTTPKEPASTTTPKEPTTTIKSA 240

QY 241 PTPKBPATTTKSAPTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTK 300  
DB 241 PTPKBPATTTKSAPTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTK 300

QY 301 PAPTPPKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTK 360  
DB 301 PAPTPPKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTK 360

QY 361 PKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTK 420  
DB 361 PKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTK 420

QY 421 TTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTK 480  
DB 421 TTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTK 480

QY 481 ETAPTPPKLTPPTPEKLAFTTPEKAPTTPEELAPTTPEELAPTTPEELAPTTPEELAP 540  
DB 481 ETAPTPPKLTPPTPEKLAFTTPEKAPTTPEELAPTTPEELAPTTPEELAPTTPEELAP 540

QY 541 NTPKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTK 600  
DB 541 NTPKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTK 600

QY 601 PTTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTK 660  
DB 601 PTTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTK 660

QY 661 KKPAPKELAPTTTKGPTSTTSKAPATTTPKETAPTTPKETAPTTPKETAPTTPKETAP 720  
DB 661 KKPAPKELAPTTTKGPTSTTSKAPATTTPKETAPTTPKETAPTTPKETAPTTPKETAP 720

QY 721 SEVSTPTTKBPATTTIHKSPDESTPELSABPTPKALENSPKBPAGVPTTKTAAKPEMTIT 780  
DB 721 SEVSTPTTKBPATTTIHKSPDESTPELSABPTPKALENSPKBPAGVPTTKTAAKPEMTIT 780

QY 781 AKDKTTERDLRTTPETTTAAPKMTKETATTTTEKTTESKITATTTQVSTTTQDTTPPKIT 840  
DB 781 AKDKTTERDLRTTPETTTAAPKMTKETATTTTEKTTESKITATTTQVSTTTQDTTPPKIT 840

QY 841 TLKTTTLAPKVTTTKITTTTEIMNKPEETAAPKDRATNSKATTPKPKPTKAPKPTST 900  
DB 841 TLKTTTLAPKVTTTKITTTTEIMNKPEETAAPKDRATNSKATTPKPKPTKAPKPTST 900

QY 901 KKPATMPRVKPKTTTPTRKMTSTWPELNPTSRABEAMLOTTTPRPNQTPNSKLVEVNPKS 960  
DB 901 KKPATMPRVKPKTTTPTRKMTSTWPELNPTSRABEAMLOTTTPRPNQTPNSKLVEVNPKS 960

QY 961 EDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS 1006

DB 961 EDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS 1006

RESULT 2

US-07-757-022B-44  
Sequence 44, Application US/07757022B  
Patent No. 6433142  
GENERAL INFORMATION:  
APPLICANT: Gesner, Thomas G.  
APPLICANT: Clark, Stephen C.  
APPLICANT: Turner, Katherine  
APPLICANT: Hewick, Rodney M.  
TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
NUMBER OF SEQUENCES: 143  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/757,022B  
FILING DATE: 19910910  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Cserr, Luann  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 5190  
TELEPHONE: (617)876-1170  
TELEFAX: (617)876-5851  
INFORMATION FOR SEQ ID NO: 44:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1270 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-757-022B-44

Query Match 100.0%; Score 5373; DB 4; Length 1270;  
Best Local Similarity 100.0%; Pred. No. 2e-167;  
Matches 1006; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQELSCKGRCFESFERGRCDCDAQCKKYDKCCPDYE 60  
DB 1 MAWKTLPIYLLLLSVFVIQVSSQELSCKGRCFESFERGRCDCDAQCKKYDKCCPDYE 60

QY 61 SFCAEVKDNKNRTKKKPTKPPVDEAGSLDNGDFKVTTPDTSTTQHKNVSTSPKITT 120  
DB 61 SFCAEVKDNKNRTKKKPTKPPVDEAGSLDNGDFKVTTPDTSTTQHKNVSTSPKITT 120

QY 121 AKPINRPSLPNSDTSKETSITVKNKETTVEKTTTNNKQTSIDGKEKTTSAKETQSIE 180  
DB 121 AKPINRPSLPNSDTSKETSITVKNKETTVEKTTTNNKQTSIDGKEKTTSAKETQSIE 180



Db 421 PSPTPKBPAPTTTSAPTTKEPAPTTSAPTTPKSPSTTTTKEPAPTTPKEPAPTTP 480  
Qy 440 KKPAPTTPKEPAPTTPKEPAPTTPKBPAPTAKBPAPTTPKEAPTTPKLTPTTPEKLA 499  
Db 481 KKPAPTTPKEPAPTTPKEPAPTTPKBPAPTAKBPAPTTPKEAPTTPKLTPTTPEKLA 540  
Qy 500 PTTPEKAPTTPPEELAPTTPEPTPTTPEEPAPTTPKAAAPTTPKEPAPTTPKEPAPTTP 559  
Db 541 PTTPEKAPTTPPEELAPTTPEPTPTTPEEPAPTTPKAAAPTTPKEPAPTTPKEPAPTTP 600  
Qy 560 KKPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKAPKELAPTTTKEPTSTTSDEKAPTTP 619  
Db 601 KKPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKAPKELAPTTTKEPTSTTSDEKAPTTP 660  
Qy 620 PKGTAPTTPKEPAPTTPKBPAPTTPKGTAPTTLKEPAPTTPKAPKELAPTTTKEPTST 679  
Db 661 PKGTAPTTPKEPAPTTPKBPAPTTPKGTAPTTLKEPAPTTPKAPKELAPTTTKEPTST 720  
Qy 680 TSDKAPTTPKEPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTP 739  
Db 721 TSDKAPTTPKEPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTP 780  
Qy 740 DESTPELSABPTPKALENSPKBPVPTTKTPAATKPEMTTTAKDKTTERDLRTPPTTTA 799  
Db 781 DESTPELSABPTPKALENSPKBPVPTTKTPAATKPEMTTTAKDKTTERDLRTPPTTTA 840  
Qy 800 APRMTKETATTEKTTESKITATTTQVSTTTQDTTPPKLTTLKTTTLAPKVTTKKTTIT 859  
Db 841 APRMTKETATTEKTTESKITATTTQVSTTTQDTTPPKLTTLKTTTLAPKVTTKKTTIT 900  
Qy 860 TTEIMNKPBEAKPKDRATNSKATPKBPQKPTKAPKPTSTKPKPMRVRKPKTTPTPR 919  
Db 901 TTEIMNKPBEAKPKDRATNSKATPKBPQKPTKAPKPTSTKPKPMRVRKPKTTPTPR 960  
Qy 920 KMTSTWPELNPTSRIAEAMLOTTTRNQTPNSKLVEVNPKSEDAGGAEGETPHMLLRPHV 979  
Db 961 KMTSTWPELNPTSRIAEAMLOTTTRNQTPNSKLVEVNPKSEDAGGAEGETPHMLLRPHV 1020  
Qy 980 FMEVPTPDMYLRVNPQGHIIINPMLS 1006  
Db 1021 FMEVPTPDMYLRVNPQGHIIINPMLS 1047

## RESULT 4

US-07-757-022B-58  
; Sequence 58, Application US/07757022B  
; Patent No. 6433142

## GENERAL INFORMATION:

; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 Cambridgepark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140

## COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserik, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 58:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1049 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-07-757-022B-58

Query Match 99.7%; Score 5358.7; DB 4; Length 1049;

Best Local Similarity 95.9%; Pred. No. 4.5e-167;

Matches 1006; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

Qy 1 MAWKTLPIVLLLLSVFVVIQVSSQBLSCGRCFESFERGRCDCDAOCKYDKCCPDYE 60  
Db 1 MAWKTLPIVLLLLSVFVVIQVSSQBLSCGRCFESFERGRCDCDAOCKYDKCCPDYE 60  
Qy 61 SPCAE-----VKNKKNTKKK 77  
Db 61 SPCAEHSVSEHQESSSSSSSSSIWKIKSSKNSAANRELQKLKVKDKNKNRTKKK 120  
Qy 78 PTPKPPVVDVAGSLDNGDFKVTTPDTSTTQHKNVSTSPKITTAKPINRPSLPNDSDT 137  
Db 121 PTPKPPVVDVAGSLDNGDFKVTTPDTSTTQHKNVSTSPKITTAKPINRPSLPNDSDT 180  
Qy 138 KETSLTVNKEVTVETKETTNNKQSTGDKKETTSAKETSIEKTSADKLAPTSKVLAKP 197  
Db 181 KETSLTVNKEVTVETKETTNNKQSTGDKKETTSAKETSIEKTSADKLAPTSKVLAKP 240  
Qy 198 TPKAETTTKGPALTTPKEPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTPKSAPT 257  
Db 241 TPKAETTTKGPALTTPKEPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTPKSAPT 300  
Qy 258 TPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 317  
Db 301 TPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 360  
Qy 318 APTTPKEPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 377  
Db 361 APTTPKEPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 420  
Qy 378 KEPSPTTPKEPAPTTPKSAPTTTKSAPTTTKSAPTTTKSAPTTTKSAPTTTKSAPTTP 437  
Db 421 KEPSPTTPKEPAPTTPKSAPTTTKSAPTTTKSAPTTTKSAPTTTKSAPTTTKSAPTTP 480  
Qy 438 TPKKPAPTTPKEPAPTTPKEPAPTTPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAP 497  
Db 481 TPKKPAPTTPKEPAPTTPKEPAPTTPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAP 540  
Qy 498 LAPTPKEPAPTTPPEELAPTTPEEPPTTPEEPAPTTPKAAAPTTPKEPAPTTPKEPAPT 557  
Db 541 LAPTPKEPAPTTPPEELAPTTPEEPPTTPEEPAPTTPKAAAPTTPKEPAPTTPKEPAPT 600  
Qy 558 TPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKAPKAPKAPKAPKAPKAPKAP 617  
Db 601 TPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKAPKAPKAPKAPKAPKAPKAP 660  
Qy 618 TTPKGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKAPKAPKAPKAP 677





QY 858 ITTTEIMNKPBTAKPKDRATNSKATTPKQKPTKAPKKSTSTKKPKTMPRVKPKTPT 917  
DB 901 ITTTEIMNKPBTAKPKDRATNSKATTPKQKPTKAPKKSTSTKKPKTMPRVKPKTPT 960  
QY 918 PRKMTSTMPNLNPTSRIAEAMLOTTTRPNQTPNSKLVEVNPXSEDAGGAEGTPEHMLLRP 977  
DB 961 PRKMTSTMPNLNPTSRIAEAMLOTTTRPNQTPNSKLVEVNPXSEDAGGAEGTPEHMLLRP 1020  
QY 978 HVFMEVPTDMDYLPRVNPQGIINPMLS 1006  
DB 1021 HVFMEVPTDMDYLPRVNPQGIINPMLS 1049

RESULT 6  
US-07-757-022B-46  
; Sequence 46, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA: US 07/546,114  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA: US 07/390,901  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cseri, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 46:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1320 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-07-757-022B-46

Query Match 99.7%; Score 5358; DB 4; Length 1320;  
Best Local Similarity 95.3%; Pred. No. 6.3e-167;  
Matches 1006; Conservative 0; Mismatches 0; Indels 50; Gaps 1;  
QY 1 MAWKTLPIYLLLSVFVIQVSSQELSCGKRCFSEFERGECDCDAQCKKYDKCCPDYE 60

DB 1 MAWKTLPIYLLLSVFVIQVSSQELSCGKRCFSEFERGECDCDAQCKKYDKCCPDYE 60  
QY 61 SFCA-----EYKONK 70  
DB 61 SFCAVHNPTSPSSKAPPAGSQTIKSTIKRSPKPNKKTKKVIESEBITEVONK 120  
QY 71 KNRTKKKTPKPPVVDEAGSLDNGDFKVTTPDSTTQHNKVVSTSPKITTAKPINRPSL 130  
DB 121 KNRTKKKTPKPPVVDEAGSLDNGDFKVTTPDSTTQHNKVVSTSPKITTAKPINRPSL 180  
QY 131 PPNSTSKETSLSLVNKKETTVETKETTINKQTSIDGKEKTTSAKETOSLEKTSADLAPT 190  
DB 181 PPNSTSKETSLSLVNKKETTVETKETTINKQTSIDGKEKTTSAKETOSLEKTSADLAPT 240  
QY 191 SKVLAKPTPKAETTTKGPALITPKPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPT 250  
DB 241 SKVLAKPTPKAETTTKGPALITPKPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPT 300  
QY 251 TTKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKKA 310  
DB 301 TTKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKKA 360  
QY 311 PTTPEPAPTTPKEPTPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPK 370  
DB 361 PTTPEPAPTTPKEPTPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPK 420  
QY 371 EPAPTTTKEPSPTTPKEPAPTTPKSAPTTTKSAPTTTKSAPTTTKSAPTTTKSAPTTTK 430  
DB 421 EPAPTTTKEPSPTTPKEPAPTTPKSAPTTTKSAPTTTKSAPTTTKSAPTTTKSAPTTTK 480  
QY 431 PKEPAPTTPKKAAPTTPKEPAPTTPKEPAPTTPKKAAPTTPKKAAPTTPKKAAPTTPK 490  
DB 481 PKEPAPTTPKKAAPTTPKEPAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPK 540  
QY 491 TPTTEKLAPTTPKEKAPTTPBELAPTTPKEPAPTTPKEPAPTTPKKAAPTTPKKAAPTTP 550  
DB 541 TPTTEKLAPTTPKEKAPTTPBELAPTTPKEPAPTTPKEPAPTTPKKAAPTTPKKAAPTTP 600  
QY 551 PKEPAPTTPKEPAPTTPKETAPTTPKGAAPTTLKAPAPTTPKKAAPTTPKKAAPTTPK 610  
DB 601 PKEPAPTTPKEPAPTTPKETAPTTPKGAAPTTLKAPAPTTPKKAAPTTPKKAAPTTPK 660  
QY 611 TSDKAPTTPKGAAPTTPKEPAPTTPKEPAPTTPKGAAPTTLKAPAPTTPKKAAPTTPK 670  
DB 661 TSDKAPTTPKGAAPTTPKEPAPTTPKEPAPTTPKGAAPTTLKAPAPTTPKKAAPTTPK 720  
QY 671 TTTKGPTSTSDKAPTTPKETAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTP 730  
DB 721 TTTKGPTSTSDKAPTTPKETAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTP 780  
QY 731 EPTTIHKSDESTEPKSAEPTPKALENSKPEKGVPTTKTAAATKPEMTTTAKDKTTERDL 790  
DB 781 EPTTIHKSDESTEPKSAEPTPKALENSKPEKGVPTTKTAAATKPEMTTTAKDKTTERDL 840  
QY 791 RTTPETTTAAPKMTKETAATTEKTTESKITAATTTQVSTSTTTQDTTPPKITLTKTILAPK 850  
DB 841 RTTPETTTAAPKMTKETAATTEKTTESKITAATTTQVSTSTTTQDTTPPKITLTKTILAPK 900  
QY 851 VTTTKTITTTTEIMNKPBTAKPKDRATNSKATTPKQKPTKAPKKPTSTKKPKTMPVR 910  
DB 901 VTTTKTITTTTEIMNKPBTAKPKDRATNSKATTPKQKPTKAPKKPTSTKKPKTMPVR 960  
QY 911 KPMTTPTPKMTSTMPNLNPTSRIAEAMLOTTTRPNQTPNSKLVEVNPXSEDAGGAEGT 970  
DB 961 KPMTTPTPKMTSTMPNLNPTSRIAEAMLOTTTRPNQTPNSKLVEVNPXSEDAGGAEGT 1020  
QY 971 PHMLLRPHVEMPEVTPDMDYLPRVNPQGIINPMLS 1006  
DB 1021 PHMLLRPHVEMPEVTPDMDYLPRVNPQGIINPMLS 1056



ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/757,022B  
FILING DATE: 19910910  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Cserr, Luann  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 5190  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)876-1170  
TELEFAX: (617)876-5851  
INFORMATION FOR SEQ ID NO: 52:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1363 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-757-022B-52

Query Match 99.6%; Score 5353.7; DB 4; Length 1363;  
Best Local Similarity 91.5%; Pred. No. 9.1e-167;  
Matches 100%; Conservative 0; Mismatches 0; Indels 93; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQELSCKGRGCFESFERGECDCDAQCKYDKCCPDYE 60  
Db 1 MAWKTLPIYLLLLSVFVIQVSSQELSCKGRGCFESFERGECDCDAQCKYDKCCPDYE 60  
QY 61 SFCAE----- 65  
Db 61 SFCAEVHNPTSPPSKKAPPSPGASQTIKSTTKRSPKPNKKTKKVIESEIEBHSVS 120  
QY 66 -----VKDNKNRTKKKTPKPPVWDE 87  
Db 121 ENQESSSSSSSSSTIWKIKSSKNSAANRELQKKLVKDNKNRTKKKTPKPPVWDE 180  
QY 88 AGSGLDNGDKVTPDTSSTTHQNKVSTSPKITTAKPINRPSLPNSDTSKETSLSVKNKE 147  
Db 181 AGSGLDNGDKVTPDTSSTTHQNKVSTSPKITTAKPINRPSLPNSDTSKETSLSVKNKE 240  
QY 148 TTIVETKETTNNKQSTPDGKEKTTSAKETQSIBKTSKAKDLAPTSKVLAKTPKAEATTIKG 207  
Db 241 TTIVETKETTNNKQSTPDGKEKTTSAKETQSIBKTSKAKDLAPTSKVLAKTPKAEATTIKG 300  
QY 208 PALTTKPEPTTPKPEASTTKPEPTTIKSAPTTPKPEAPTTTKSAPTTPKPEAPTTT 267  
Db 301 PALTTKPEPTTPKPEASTTKPEPTTIKSAPTTPKPEAPTTTKSAPTTPKPEAPTTT 360  
QY 268 KEPAPTTKPEAPTTTKSAPTTPKPEAPTTTKKPAPTTPKPEAPTTTKPEPTTPKEPTP 327  
Db 361 KEPAPTTKPEAPTTTKSAPTTPKPEAPTTTKKPAPTTPKPEAPTTTKPEPTTPKEPTP 420  
QY 328 TTPKEAPTTKPEAPTTKPEAPTTAPKKPAPTTPKPEAPTTTKPEAPTTTKESPTTPKPE 387  
Db 421 TTPKEAPTTKPEAPTTKPEAPTTAPKKPAPTTPKPEAPTTTKPEAPTTTKESPTTPKPE 480

QY 388 PAPTTTKGAPTTTKPEAPTTTKSAPTTPKPESEPTTKPEAPTTTKPEAPTTTKPKKPAPTTP 447  
Db 481 PAPTTTKGAPTTTKPEAPTTTKSAPTTPKPESEPTTKPEAPTTTKPEAPTTTKPKKPAPTTP 540  
QY 448 KEPAPTTKPEAPTTTKKPAPTAPKPEAPTTTKETAPTTPKLTPTTPEKLAPTTPEKPA 507  
Db 541 KEPAPTTKPEAPTTTKKPAPTAPKPEAPTTTKETAPTTPKLTPTTPEKLAPTTPEKPA 600  
QY 508 PTTPEELAPTTPEEPTTPKPEAPTTTKAAANPTPKPEAPTTTKPEAPTTTKPEAPTTTK 567  
Db 601 PTTPEELAPTTPEEPTTPKPEAPTTTKAAANPTPKPEAPTTTKPEAPTTTKPEAPTTTK 660  
QY 568 KETAPTTKGTAPTTLKPEAPTTPKKPAKELAPTTTKETPTSTTSKPAPTTKGTAPTT 627  
Db 661 KETAPTTKGTAPTTLKPEAPTTPKKPAKELAPTTTKETPTSTTSKPAPTTKGTAPTT 720  
QY 628 PKPEAPTTKPEAPTTPKGTAPTTLKPEAPTTPKKPAKELAPTTTKGTSTTSKPAPT 687  
Db 721 PKPEAPTTKPEAPTTPKGTAPTTLKPEAPTTPKKPAKELAPTTTKGTSTTSKPAPT 780  
QY 688 TPKETAPTTKPEAPTTPKKPAPTTPETPPPTTSEVSTPTTKEPTTIHKSPDESTEP 747  
Db 781 TPKETAPTTKPEAPTTPKKPAPTTPETPPPTTSEVSTPTTKEPTTIHKSPDESTEP 840  
QY 748 ABPTPKALENSPKPGVPTTKTAAATKPEMTTTAKDKITERDLRTTPTTTAAPKMTKET 807  
Db 841 ABPTPKALENSPKPGVPTTKTAAATKPEMTTTAKDKITERDLRTTPTTTAAPKMTKET 900  
QY 808 ATTTEKTESKITATTQVTSSTTQDTPPKITLTKTTLAPKVITTKITITTEIMNKP 867  
Db 901 ATTTEKTESKITATTQVTSSTTQDTPPKITLTKTTLAPKVITTKITITTEIMNKP 960  
QY 868 EETAKPKDEATNSKATTPKPKQPKTAPKPKPTSTKKPKTMRVRKPKTTPTRKMTSTMPE 927  
Db 961 EETAKPKDEATNSKATTPKPKQPKTAPKPKPTSTKKPKTMRVRKPKTTPTRKMTSTMPE 1020  
QY 928 LNPTSRIBAMLQTTTTRPNQTPNSKLVEVNPKSEDAGGAEGETHMLLRPHVFMPEVTPD 987  
Db 1021 LNPTSRIBAMLQTTTTRPNQTPNSKLVEVNPKSEDAGGAEGETHMLLRPHVFMPEVTPD 1080  
QY 988 MDYLPRVFNQGIINPMLS 1006  
Db 1081 MDYLPRVFNQGIINPMLS 1099  
RESULT 9  
US-10-164-595-58  
; Sequence 58, Application US/10164595  
; Patent No. 6657054  
; GENERAL INFORMATION:  
; APPLICANT: Origene Technologies, Inc  
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides  
; FILE REFERENCE: IU 103 R1  
; CURRENT APPLICATION NUMBER: US/10/164,595  
; CURRENT FILING DATE: 2002-06-10  
; NUMBER OF SEQ ID NOS: 80  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 58  
; LENGTH: 1320  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-164-595-58  
Query Match 99.6%; Score 5349; DB 4; Length 1320;  
Best Local Similarity 95.1%; Pred. No. 1.2e-166;  
Matches 100%; Conservative 0; Mismatches 2; Indels 50; Gaps 1;  
QY 1 MAWKTLPIYLLLLSVFVIQVSSQELSCKGRGCFESFERGECDCDAQCKYDKCCPDYE 60  
Db 1 MAWKTLPIYLLLLSVFVIQVSSQELSCKGRGCFESFERGECDCDAQCKYDKCCPDYE 60  
QY 61 SFCAE-----EYKDNK 70  
|||||

61	Db	SFCAEVHNPTSPFPSSKAPPSGASQTTIKSTTKRGSPKPNKKTKKVIJESSEITEVQDNK	120
71	Qy	KNRTKKKPTPPVPPVDEAGSLDNGDFKVTTPDTSITQHKNKVSTSPGKITTAKPINPRPSL	130
121	Db	KNRTKKKPTPPVPPVDEAGSLDNGDFKVTTPDTSITQHKNKVSTSPKITTAKPINPRPSL	180
131	Qy	PPNGDTSKETSIVNKEITTVETKETITTNKQTSITDCKEKTTSAKETSQSIKTSKADLAPT	190
181	Db	PPNGDTSKETSIVNKEITTVETKETITTNKQTSITDCKEKTTSAKETSQSIKTSKADLAPT	240
191	Qy	SKVLAKPTPKAEITTTKGPAITTPKPEPASTTPKPEPTPTTIKSAPITTPKPEAPT	250
241	Db	SKVLAKPTPKAEITTTKGPAITTPKPEPASTTPKPEPTPTTIKSAPITTPKPEAPT	300
251	Qy	TTKSAPITPKBPAPITTTKPEAPITTPKPEAPITTTKPEAPITTTKSAPITTPKPEAPTTPPKPA	310
301	Db	TTKSAPITPKBPAPITTTKPEAPITTPKPEAPITTTKPEAPITTTKSAITTPKPEAPITTPKPA	360
311	Qy	PTTPKPEAPITTPKPEPTPTTPKPEAPITTPKPEAPITTPKPEAPITTPKPEAPITTPK	370
361	Db	PTTPKPEAPITTPKPEPTPTTPKPEAPITTPKPEAPITTPKPEAPITTPKPEAPITTPK	420
371	Qy	EPAPITTTKEPSPTTPKBPAPITTTKSAPITTTKPEAPITTTKSAPITTPKBPSPITTKPEAPT	430
421	Db	EPAPITTTKEPSPTTPKBPAPITTTKSAPITTTKPEAPITTTKSAPITTPKBPSPITTKPEAPT	480
431	Qy	PKPEAPITTPKBPAPITTPKPEAPITTTKKPAPTAPKPEAPITTPKETAITTPKKL	490
481	Db	PKPEAPITTPKBPAPITTPKPEAPITTTKKPAPTTPKPEAPITTPKETAITTPKKL	540
491	Qy	TPTTPEKLAITTPPKBPAPITTPPEELAPITTPPEPTPTTPPEEPAPITTPKAAAPNTPKPEAPT	550
541	Db	TPTTPEKLAITTPPKBPAPITTPPEELAPITTPPEPTPTTPPEEPAPITTPKAAAPNTPKPEAPT	600
551	Qy	PKPEAPITTPKPEAPITTPKETAITTPKGTAPITTLKPEAPITTPKKPAPKELAITTTKEPTST	610
601	Db	PKPEAPITTPKPEAPITTPKETAITTPKGTAPITTLKPEAPITTPKKPAPKELAITTTKEPTST	660
611	Qy	TSOKPAPITTPKGTAPITTPKBPAPITTPKPEAPITTPKGTAPITTLKPEAPITTPKKPAPKELAP	670
661	Db	TSOKPAPITTPKGTAPITTPKBPAPITTPKPEAPITTPKGTAPITTLKPEAPITTPKKPAPKELAP	720
671	Qy	TTTKGPTSTTSOKPAPITTPKETAITTPKPEAPITTPKPEAPITTPPEPTPTTSEVSTPTTK	730
721	Db	TTTKGPTSTTSOKPAPITTPKETAITTPKPEAPITTPKPEAPITTPPEPTPTTSEVSTPTTK	780
731	Qy	EPTTIHKSPDESTPELSAETTPKALENSKPEPGVPTTKTAATKPEMTTTAKDKTTBRDL	790
781	Db	EPTTIHKSPDESTPELSAETTPKALENSKPEPGVPTTKTAATKPEMTTTAKDKTTBRDL	840
791	Qy	RTTPETTTAAPKMTKETATTTTEKTTESKIIATTTQVITSTTTQDPTTPPKITTLKTTTLAPK	850
841	Db	RTTPETTTAAPKMTKETATTTTEKTTESKIIATTTQVITSTTTQDPTTPPKITTLKTTTLAPK	900
851	Qy	VTTTKTITTTTEIIMNKPEETAKPDRATNSKATTPPKOKPTKAPKQTSITKKPKTMRVR	910
901	Db	VTTTKTITTTTEIIMNKPEETAKPDRATNSKATTPPKOKPTKAPKQTSITKKPKTMRVR	960
911	Qy	KPKTTPTPKMTSTMPELNPTSIAEAMLOTTTPRPNQTPNSKLVEVNPKSGEDAGGABGET	970
961	Db	KPKTTPTPKMTSTMPELNPTSIAEAMLOTTTPRPNQTPNSKLVEVNPKSGEDAGGABGET	1020
971	Qy	PHMLLRPHVFMPEVTPDMDYLPRVNPGLIINPMLS	1006
1021	Db	PHMLLRPHVFMPEVTPDMDYLPRVNPGLIINPMLS	1056

## RESULT 10

RESOL TO  
US-07-757-022B-48

; Sequence 48, Application US/07757022B

; Patent No. 6433142

GENERAL INFORMATION:

; APPLICANT: Gesner, Thomas G.

/ APPLICANT: Clark, Stephen C.  
 / APPLICANT: Turner, Katherine  
 / APPLICANT: Hewick, Rodney M.  
 / TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
 / NUMBER OF SEQUENCES: 143  
 / CORRESPONDENCE ADDRESS:  
 / ADDRESSEE: Genetics Institute, Inc.  
 / STREET: 87 Cambridgepark Drive  
 / CITY: Cambridge  
 / STATE: Massachusetts  
 / COUNTRY: U.S.A.  
 / ZIP: 02140  
 / COMPUTER READABLE FORM:  
 / MEDIUM TYPE: Floppy disk  
 / COMPUTER: IBM PC compatible  
 / OPERATING SYSTEM: PC-DOS/MS-DOS  
 / SOFTWARE: Patent In Release #1.0, Version #1.25  
 / CURRENT APPLICATION DATA:  
 / APPLICATION NUMBER: US 07/757,022B  
 / FILING DATE: 19910910  
 / CLASSIFICATION: 530  
 / PRIOR APPLICATION DATA:  
 / APPLICATION NUMBER: US 07/643,502  
 / FILING DATE: 18-JAN-1991  
 / PRIOR APPLICATION DATA:  
 / APPLICATION NUMBER: US 07/546,114  
 / FILING DATE: 29-JUN-1990  
 / PRIOR APPLICATION DATA:  
 / APPLICATION NUMBER: US 07/457,196  
 / FILING DATE: 29-DEC-1989  
 / PRIOR APPLICATION DATA:  
 / APPLICATION NUMBER: US 07/390,901  
 / FILING DATE: 08-AUG-1989  
 / ATTORNEY/AGENT INFORMATION:  
 / NAME: Cserri, Luann  
 / REGISTRATION NUMBER: 31,822  
 / REFERENCE/DOCKET NUMBER: GI 5190  
 / TELECOMMUNICATION INFORMATION:  
 / TELEPHONE: (617)876-1170  
 / TELEFAX: (617)876-5851  
 / INFORMATION FOR SEQ ID NO: 48:  
 / SEQUENCE CHARACTERISTICS:  
 / LENGTH: 1354 amino acids  
 / TYPE: AMINO ACID  
 / TOPOLOGY: linear  
 / MOLECULE TYPE: protein  
 / US-07-757-022B-48

Query Match	99.5%;	Score 5344.6;	DB 4;	Length 1354;
Best Local Similarity	92.3%;	Pred. No. 1.8e-166;		
Matches 1006;	Conservative 0;	Mismatches 0;	Indels 84;	Gaps 2
Qy	1	MAWKTLPIYLILLLSVRFVIQQVSSQ-----	25	
Dd	1	MAWKTLPIYLILLLSVRFVIQQVSSQLSSCAGRCGEGYSRDATCNCYDNCQHVMCCPDF	60	
Qy	26	-----ELSCCKGRCFSEFREGRECDDAQCKKYDKCCPDYBSFCAE-----	65	
Dd	61	KRVCTAELSCCKGRCFSEFREGRECDDAQCKKYDKCCPDYBSFCAEHHSVSNEGESSSSS	120	
Qy	66	-----VKDNKKNRTKTKKPTPKPPVPVDGAGSLNDGD	96	
Dd	121	SSSSSSSTIWKIKSSKNSSAAARELQKKLVKDKNKKNTKTKKPTPKPPVPVDGAGSLNDGD	180	
Qy	97	FKVTTPDTSTOHNKVYSTSPKIITAKPINRPSLPNSDTSKETSLTVNKETTIVETKETT	156	
Dd	181	FKVTTPDTSTOHNKVYSTSPKIITAKEINRPSLPNSDTSKETSLTVNKETTIVETKETT	240	
Qy	157	TTNKQTSTDGEKITTSACKETQSLEKTSASAKDLAPTSKVLAKPTPKAEFTTTKGPALTTTKEP	216	
Dd	241	TTNKQTSTDGEKITTSACKETQSLEKTSASAKDLAPTSKVLAKPTPKAEFTTTKGPALTTTKEP	300	
Qy	217	TPPTPKPEASTTKEPTPTTIKGAPTTPKEPAFTTTKSAPTTPKEPATPTTKSPAPTTPK	276	











Db 781 TAPTLKEPAPPTPKKPAKELAPTTTKGPTSTTSKPAPTTPKETAPPTPKPAPPTPK 840  
QY 707 KPAPTTPPTPTSTVSTPTTTTKEPTTIHKSPDESPKALNSPKKPGVPT 766  
Db 841 KPAPTTPPTPTSTVSTPTTTTKEPTTIHKSPDESPKALNSPKKPGVPT 900  
QY 767 TKTPAATKPEMTTAKDKTTERDLRTPTTAAAPKMKETATTTTEKTTESKITATTQV 826  
Db 901 TKTPAATKPEMTTAKDKTTERDLRTPTTAAAPKMKETATTTTEKTTESKITATTQV 960  
QY 827 TSTTTQDTTPFKITTLTKTTLAPKVTTKKVIETSEBELTEHVSVENQESSSSSSSTI 886  
Db 961 TSTTTQDTTPFKITTLTKTTLAPKVTTKKVIETSEBELTEHVSVENQESSSSSSSTI 1020  
QY 887 POKPTKAPKPTSTKPKMVRKPKTTPTRKMTSTMPKMTSRIAEAMLTQTTTRN 946  
Db 1021 POKPTKAPKPTSTKPKMVRKPKTTPTRKMTSTMPKMTSRIAEAMLTQTTTRN 1080  
QY 947 QTPNSKLVEVNPKESEDAGAGETPHMLLRPHVFMPEVTPDMDYLRVFNQGIINPMLS 1006  
Db 1081 QTPNSKLVEVNPKESEDAGAGETPHMLLRPHVFMPEVTPDMDYLRVFNQGIINPMLS 1140

RESULT 14  
US-07-757-022B-62  
; Sequence 62, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ceert, Luan  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 62:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1404 amino acids  
; TYPE: AMINO ACID

; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-07-757-022B-62  
Query Match 99.4%; Score 5339.6; DB 4; Length 1404;  
Best Local Similarity 88.2%; Pred. No. 2.7e-166;  
Matches 1006; Conservative 0; Mismatches 0; Indels 134; Gaps 2;  
QY 1 MAWKTIPIYLLLLSVFVIQOVSSQ-----25  
Db 1 MAWKTLPIVLLLLSVFVIQOVSSQDLSSCAGRCGEGYSDATCNCNDYNCQHMECCPDF 60  
QY 26 -----ELSCKGRCFSEFERGECDCDAQCKKYDKCCPDYESFCAE-----65  
Db 61 KRVTAEELCKGRCFSEFERGECDCDAQCKKYDKCCPDYESFCAEVHNPTSPSSKKAP 120  
QY 66 -----65  
Db 121 PPSGASQTIKSTTKRSPKPNKKTKKVIETSEBELTEHVSVENQESSSSSSSSSTI 180  
QY 66 -----VKONKXRTKKKPTPKPPVVDVDEAGSLDNGDFKVTPTDST 106  
Db 181 KIKSSKNSAANRELQKKLVKONKXRTKKKPTPKPPVVDVDEAGSLDNGDFKVTPTDST 240  
QY 107 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVNKETTVEKETTINKQTSIDG 166  
Db 241 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVNKETTVEKETTINKQTSIDG 300  
QY 167 KEKTTISAKETQSIKTSKDLAPTSSKVLAKPTPKAETTTKGPALTPPKPTPTPKEPAS 226  
Db 301 KEKTTISAKETQSIKTSKDLAPTSSKVLAKPTPKAETTTKGPALTPPKPTPTPKEPAS 360  
QY 227 TTPKPTPTTIKSAPTTPKEPAPTTTKSAPTTKSGEAPTTTKEPAPTTTKEPAPTTTKEP 286  
Db 361 TTPKPTPTTIKSAPTTPKEPAPTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 420  
QY 287 APTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 346  
Db 421 APTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 480  
QY 347 EPAPTAPKKPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 406  
Db 481 EPAPTAPKKPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 540  
QY 407 TTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 466  
Db 541 TTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 600  
QY 467 APTAPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 526  
Db 601 APTAPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 660  
QY 527 PEEPAPTTPKAAAPNTPKPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 586  
Db 661 PEEPAPTTPKAAAPNTPKPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 720  
QY 587 APTPKKAPKELAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 646  
Db 721 APTPKKAPKELAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 780  
QY 647 TAPTTTKEPAPTTPKKAPKELAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 706  
Db 781 TAPTTTKEPAPTTPKKAPKELAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 840  
QY 707 KPAPTTPPTPTSTVSTPTTTTKEPTTIHKSPDESPKALNSPKKPGVPT 766  
Db 841 KPAPTTPPTPTSTVSTPTTTTKEPTTIHKSPDESPKALNSPKKPGVPT 900  
QY 767 TKTPAATKPEMTTAKDKTTERDLRTPTTAAAPKMKETATTTTEKTTESKITATTQV 826  
Db 901 TKTPAATKPEMTTAKDKTTERDLRTPTTAAAPKMKETATTTTEKTTESKITATTQV 960  
QY 827 TSTTTQDTTPFKITTLTKTTLAPKVTTKKVIETSEBELTEHVSVENQESSSSSSSTI 886

```
Db 961 TSTTTQDTPPKITLTKLTLAPKVTITTKLITITTEIMNKPEETAKPKDRATNSKATTPK 1020
QY 887 POKPTKAPKPTSTKKPKTMPVRKPKTTPTPRKMTSTMPENLPTSRIAEAMLQTTTRN 946
Db 1021 POKPTKAPKPTSTKKPKTMPVRKPKTTPTPRKMTSTMPENLPTSRIAEAMLQTTTRN 1080
QY 947 QTPNSKLVEVNPKSEDAGABGETPHMLLRPHVFMPEVTPDMDYLRVFNQGIINPMLS 1006
Db 1081 QTPNSKLVEVNPKSEDAGABGETPHMLLRPHVFMPEVTPDMDYLRVFNQGIINPMLS 1140

RESULT 15
US-09-298-970A-1
; Sequence 1, Application US/09298970A
; Patent No. 6743774
; GENERAL INFORMATION:
; APPLICANT: Jav, Gregory D.
; TITLE OF INVENTION: TRIBONECTINS
; FILE REFERENCE: 21486-026
; CURRENT APPLICATION NUMBER: US/09/298,970A
; CURRENT FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1404
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-298-970A-1

Query Match 99.4%; Score 5339.6; DB 4; Length 1404;
Best Local Similarity 88.2%; Pred. No. 2.7e-166;
Matches 1006; Conservative 0; Mismatches 0; Indels 134; Gaps 2;

QY 1 MAWKTLPIYLLLLSVFVIQQVSSQ----- 25
Db 1 MAWKTLPIYLLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF 60
QY 26 -----ELSCGRCFRSPRGRECDQDAQCKKYDKCCPDYESFCAE----- 65
Db 61 KRVCYTAELSCGRCFRSPRGRECDQDAQCKKYDKCCPDYESFCAEYHNPSTSPSSKKAP 120
QY 66 ----- 65
Db 121 PPSGASQTIKSTYKRSKPNKKTKKVIIESEBITBEHSVSNQESSSSSSSSSSSTIW 180
QY 66 -----VKDNKNKRTKKKTPKPPVVDVDEAGSGLDNGDFKVTTPDTST 106
Db 181 KIKSSKNSAANREBLQKKLVKDNKNKRTKKKTPKPPVVDVDEAGSGLDNGDFKVTTPDTST 240
QY 107 TQHNKUSTSPKITITAKPINRPSLPNSDTSKETSLSLVNKEITTVETKETTNNKQJSTDG 166
Db 241 TQHNKVSTSPKITITAKPINRPSLPNSDTSKETSLSLVNKEITTVETKETTNNKQJSTDG 300
QY 167 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALITTPKEPTTTPKEPAS 226
Db 301 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALITTPKEPTTTPKEPAS 360
QY 227 TTPKEPTPTTIKSAPTTKPEPAPTTTKSAPTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPE 286
Db 361 TTPKEPTPTTIKSAPTTKPEPAPTTTKSAPTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPE 420
QY 287 APPTTKSAPTTKPEPAPTTTKKAPTTKPEPAPTTKPEPTTTPKPEPTTTPKPEPAPTTTKPE 346
Db 421 APPTTKSAPTTKPEPAPTTTKKAPTTKPEPAPTTKPEPTTTPKPEPTTTPKPEPAPTTTKPE 480
QY 347 EPAPTAPKPEPAPTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKSAPTTTKSAPTTTKPEAPT 406
Db 481 EPAPTAPKPEPAPTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKSAPTTTKSAPTTTKPEAPT 540
QY 407 TTKSAPTTKPEPSTTTTKPEPAPTTKSPAPTTPKKAPPTTKGAPPTTKGAPPTTKGAPPTTKKP 466
Db 541 TTKSAPTTKPEPSTTTTKPEPAPTTKSPAPTTPKKAPPTTKGAPPTTKGAPPTTKGAPPTTKKP 600
```

```
QY 467 APTAPKEPAPTTPKETAPTTPKKLPTTTPKEKLAPTTPEKPAPTTPEELAPTTPEEPTPTT 526
Db 601 APTAPKEPAPTTPKETAPTTPKKLPTTTPKEKLAPTTPEKPAPTTPEELAPTTPEEPTPTT 660
QY 527 PPEPAPTTPKAAAPNTPKEPAPTTTPKEPAPTTPKSPAPTTPKETAPTTPKGTAPTTKLP 586
Db 661 PPEPAPTTPKAAAPNTPKEPAPTTTPKEPAPTTPKSPAPTTPKETAPTTPKGTAPTTKLP 720
QY 587 APTTPKKAPKELAPTTTTPKEPTSTSDKPAPTTPKGTAPTTKGTAPTTKPEPAPTTKSPAPTTK 646
Db 721 APTTPKKAPKELAPTTTTPKEPTSTSDKPAPTTPKGTAPTTKGTAPTTKPEPAPTTKSPAPTTK 780
QY 647 TAPTTTLKPEPAPTTPKKAPKELAPTTTTPKEPTSTSDKPAPTTPKGTAPTTKGTAPTTKPEPAPTTK 706
Db 781 TAPTTTLKPEPAPTTPKKAPKELAPTTTTPKEPTSTSDKPAPTTPKGTAPTTKGTAPTTKPEPAPTTK 840
QY 707 KPEAPTTPEPTTSEVSTPTTTPKPTTIHKSPDESTPELSAEPDTPKALENSPKPEGVPT 766
Db 841 KPEAPTTPEPTTSEVSTPTTTPKPTTIHKSPDESTPELSAEPDTPKALENSPKPEGVPT 900
QY 767 TKTPAATKPEMTTAKDKTTTERDLATTPETTTAAAPKMTKETATTTTETKTTESKITATTTOV 826
Db 901 TKTPAATKPEMTTAKDKTTTERDLATTPETTTAAAPKMTKETATTTTETKTTESKITATTTOV 960
QY 827 TSTTTQDTPPKITLTKLTLAPKVTITTKLITITTEIMNKPEETAKPKDRATNSKATTPK 886
Db 961 TSTTTQDTPPKITLTKLTLAPKVTITTKLITITTEIMNKPEETAKPKDRATNSKATTPK 1020
QY 887 POKPTKAPKPTSTKKPKTMPVRKPKTTPTPRKMTSTMPENLPTSRIAEAMLQTTTRN 946
Db 1021 POKPTKAPKPTSTKKPKTMPVRKPKTTPTPRKMTSTMPENLPTSRIAEAMLQTTTRN 1080
QY 947 QTPNSKLVEVNPKSEDAGABGETPHMLLRPHVFMPEVTPDMDYLRVFNQGIINPMLS 1006
Db 1081 QTPNSKLVEVNPKSEDAGABGETPHMLLRPHVFMPEVTPDMDYLRVFNQGIINPMLS 1140
```

Search completed: October 13, 2004, 11:59:09  
Job time : 25.4873 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 102.349 Seconds  
(without alignments)  
3171.696 Million cell updates/sec

Title: SEQ1-E

Perfect score: 5373

Sequence: 1 MAWKTLPIYLLLLSVFVIQ.....DMDYLPRVFNQGIINPMLS 1006

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.1

Searched: 1356558 seqs, 322682953 residues

Total number of hits satisfying chosen parameters: 1356558

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match %	Length	DB	ID	Description
1	5373	100.0	1038	13	US-10-124-557-74	Sequence 74, Appl
2	5373	100.0	1270	13	US-10-124-557-44	Sequence 44, Appl
3	5358.9	99.7	1311	13	US-10-124-557-42	Sequence 42, Appl
4	5358.7	99.7	1049	13	US-10-124-557-58	Sequence 58, Appl
5	5358.7	99.7	1313	13	US-10-124-557-142	Sequence 142, Appl
6	5358	99.7	1320	13	US-10-124-557-46	Sequence 46, Appl
7	5358	99.7	1320	13	US-10-124-557-60	Sequence 60, Appl
8	5353.7	99.6	1363	13	US-10-124-557-52	Sequence 52, Appl
9	5344.6	99.5	1354	13	US-10-124-557-48	Sequence 48, Appl
10	5343.9	99.5	1361	13	US-10-124-557-40	Sequence 40, Appl
11	5339.6	99.4	1140	13	US-10-124-557-104	Sequence 104, Appl
12	5339.6	99.4	1404	9	US-09-802-207-30	Sequence 30, Appl
13	5339.6	99.4	1404	11	US-09-897-188-1	Sequence 1, Appl
14	5339.6	99.4	1404	13	US-10-124-557-2	Sequence 2, Appl

15 5339.6 99.4 1404 13 US-10-124-557-62 Sequence 62, Appl  
16 5241.8 97.6 1022 13 US-10-124-557-84 Sequence 84, Appl  
17 5232.6 97.4 1314 13 US-10-124-557-50 Sequence 50, Appl  
18 5011 93.3 941 13 US-10-124-557-14 Sequence 14, Appl  
19 1975.4 26.2 792 9 US-09-802-207-27 Sequence 27, Appl  
20 1409.9 26.2 538 14 US-10-038-694-3 Sequence 3, Appl  
21 1335.3 24.9 5179 9 US-09-922-217-1068 Sequence 1068, Appl  
22 1335.3 24.9 5179 9 US-09-833-263-1068 Sequence 1068, Appl  
23 1335.3 24.9 5179 13 US-10-025-380-1068 Sequence 1068, Appl  
24 1335.3 24.9 5179 16 US-10-734-564-121 Sequence 121, Appl  
25 1146.1 21.3 292 16 US-10-468-910-4 Sequence 4, Appl  
26 932 17.3 1460 14 US-10-295-027-428 Sequence 428, Appl  
27 920.2 17.1 1325 9 US-09-864-761-35612 Sequence 35612, Appl  
28 892.3 16.6 1367 9 US-09-801-368-108 Sequence 108, Appl  
29 845.4 15.7 6642 14 US-10-369-493-5013 Sequence 5013, Appl  
30 844.6 15.7 3507 14 US-10-369-493-5784 Sequence 5784, Appl  
31 834.6 15.5 2090 16 US-10-408-765A-2318 Sequence 2318, Appl  
32 832.7 15.3 697 15 US-10-425-114-41545 Sequence 41545, Appl  
33 822.8 15.3 1151 10 US-09-825-751A-79 Sequence 79, Appl  
34 822 15.3 19723 15 US-10-084-846A-5 Sequence 5, Appl  
35 811.2 15.1 3256 16 US-10-408-765A-174 Sequence 174, Appl  
36 811.2 15.1 2397 15 US-10-282-122A-71232 Sequence 9, Appl  
37 806.7 15.0 3256 9 US-09-919-172-98 Sequence 71232, Appl  
38 803.4 15.0 3256 10 US-09-919-039-21 Sequence 98, Appl  
39 803.4 15.0 3256 10 US-09-996-069-10 Sequence 21, Appl  
40 798.4 14.9 1255 10 US-10-171-311-158 Sequence 158, Appl  
41 798.4 14.9 1255 14 US-10-177-293-311 Sequence 311, Appl  
42 798.4 14.9 1255 16 US-10-734-564-120 Sequence 120, Appl  
43 798.4 14.9 22152 16 US-10-715-066-5 Sequence 5, Appl  
44 793.6 14.8 5935 14 US-10-243-243A-8 Sequence 8, Appl  
45 791.9 14.7

#### ALIGNMENTS

#### RESULT 1

US-10-124-557-74  
; Sequence 74, Application US/10124557  
; Publication No. US20020137894A1  
; GENERAL INFORMATION:  
; APPLICANT: Turner, Katherine  
; Clark, Stephen C.  
; Jacobs, Kenneth  
; Hewick, Rodney M.  
; Gesner, Thomas G.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 Cambridgepark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/124.557  
FILING DATE: 16-Apr-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989







```
QY 498 LAPTPKPAATTPPELAPTTPEBPTTTPKEPAPTTPKAAAPNTPKAPATTTKEPAPT 557
Db 541 LAPTPKPAATTPPELAPTTPEBPTTTPKEPAPTTPKAAAPNTPKAPATTTKEPAPT 600
QY 558 TPKEPAPTTKEPAPTTPKGATPTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSKDPAP 617
Db 601 TPKEPAPTTKEPAPTTPKGATPTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSKDPAP 660
QY 618 TTPKGATPTTPKEPAPTTPKGATPTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSKDPAP 677
Db 661 TTPKGATPTTPKEPAPTTPKGATPTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSKDPAP 720
QY 678 STTSKDPAPTTPKGATPTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSKDPAP 737
Db 721 STTSKDPAPTTPKGATPTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSKDPAP 780
QY 738 SPDESPELSAETPKALENSPKRGVPTTKTAAATKPEMTTAKOKITRDLRTTPETT 797
Db 781 SPDESPELSAETPKALENSPKRGVPTTKTAAATKPEMTTAKOKITRDLRTTPETT 840
QY 798 TAAPKMTKETATTTKTESKITATTTQVTSSTTTQDTPPKITLTKITTLAPKVTTTKKT 857
Db 841 TAAPKMTKETATTTKTESKITATTTQVTSSTTTQDTPPKITLTKITTLAPKVTTTKKT 900
QY 858 ITTETIMNKPEETAKPKDRATNSKATTPKQKPTKAPKPTSTKPKTMRVRKPKTTPT 917
Db 901 ITTETIMNKPEETAKPKDRATNSKATTPKQKPTKAPKPTSTKPKTMRVRKPKTTPT 960
QY 918 PRKMTSTMPELNPTSRISAEAMLOTTTPRPNQTPNSKLVEVNPKSEDAGGAGETPHMLLRP 977
Db 961 PRKMTSTMPELNPTSRISAEAMLOTTTPRPNQTPNSKLVEVNPKSEDAGGAGETPHMLLRP 1020
QY 978 HVFMPPEVTEDMDYLPRVFNQGIINPMLS 1006
Db 1021 HVFMPPEVTEDMDYLPRVFNQGIINPMLS 1049

RESULT 5
US-10-124-557-142.
; Sequence 142, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
```

```
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseir, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 142:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1313 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 142:
US-10-124-557-142
```

```
Query Match 99.7%; Score 5358.7; DB 13; Length 1313;
Best Local Similarity 95.9%; Pred. No. 5,4e-134;
Matches 1006; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MAWKTLPIVILLLLSVFVIQQVSSQELSCKGRGCFESFERGREGCCDAQCKKYDKCCPDYE 60
Db 1 MAWKTLPIVILLLLSVFVIQQVSSQELSCKGRGCFESFERGREGCCDAQCKKYDKCCPDYE 60
QY 61 SPCAE-----VKDNKKNTKKK 77
Db 61 SPCAEHSHSVSEQESSSSSSSSSSSIWKIKSSKNSAANRELQKKLVKDNKKNTKKK 120
QY 78 PTPKPPVDEAGSLDNGDFKVTTPDTSSTOHNVKSTSPKITTAKINPRPSLPNSDTS 137
Db 121 PTPKPPVDEAGSLDNGDFKVTTPDTSSTOHNVKSTSPKITTAKINPRPSLPNSDTS 180
QY 138 KETSLTVNKETTVEITKTTTNKQISTDQKEKTTSAKETQSIKTSKADLAPTSKVLAKP 197
Db 181 KETSLTVNKETTVEITKTTTNKQISTDQKEKTTSAKETQSIKTSKADLAPTSKVLAKP 240
QY 198 TPKAETTTKGPALTTPKETPTTTPKEPASTTTPKEPTTTIKGAPTTPKAPATTTKSAPT 257
Db 241 TPKAETTTKGPALTTPKETPTTTPKEPASTTTPKEPTTTIKGAPTTPKAPATTTKSAPT 300
QY 258 TPKEPAPTTTKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTT 317
Db 301 TPKEPAPTTTKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTT 360
QY 318 APTTPKEPTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTT 377
Db 361 APTTPKEPTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTT 420
QY 378 KEPSPTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTT 437
Db 421 KEPSPTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTT 480
QY 438 TPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTT 497
Db 481 TPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTT 540
QY 498 LAPTTPEKAPATTPPELAPTTPEBPTTTPKEPAPTTPKAAAPNTPKAPATTTKEPAPT 557
Db 541 LAPTTPEKAPATTPPELAPTTPEBPTTTPKEPAPTTPKAAAPNTPKAPATTTKEPAPT 600
QY 558 TPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTT 617
Db 601 TPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTT 660
QY 618 TTPKGATPTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTT 677
Db 661 TTPKGATPTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTT 720
QY 678 STTSKDPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTT 737
Db 721 STTSKDPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTT 780
```

QY 738 SPDESTPBLSEPTPKALENSKPEGVPTTKTPAATKPEMTTAKDKTTERDLRTTPET 797  
Db 781 SPDESTPBLSEPTPKALENSKPEGVPTTKTPAATKPEMTTAKDKTTERDLRTTPET 840  
QY 798 TAAPKMTKETAATTTKTESKITATTTQVTSSTTQDTPPKITTLKTTTLAPKVTITTKT 857  
Db 841 TAAPKMTKETAATTTKTESKITATTTQVTSSTTQDTPPKITTLKTTTLAPKVTITTKT 900  
QY 858 ITTTEIMNKPEETAKPKORATNSKATTPKPKQKPTKAPKKPTSTKKPKMPVRKPKTTP 917  
Db 901 ITTTEIMNKPEETAKPKORATNSKATTPKPKQKPTKAPKKPTSTKKPKMPVRKPKTTP 960  
QY 918 PKMTSTMPBLNPTSGRIASAMLOTTTRPNQTPNSKLVEVNPXSEDAGGABGETPHMLLRP 977  
Db 961 PKMTSTMPBLNPTSGRIASAMLOTTTRPNQTPNSKLVEVNPXSEDAGGABGETPHMLLRP 1020  
QY 978 HVFMPVETPDMDYLPVRPNQGIINPMLS 1006  
Db 1021 HVFMPVETPDMDYLPVRPNQGIINPMLS 1049

RESULT 6

US-10-124-557-46  
; Sequence 46, Application US/10124557  
; Publication No. US20020137894A1  
; GENERAL INFORMATION:  
; APPLICANT: Turner, Katherine  
; Clark, Stephen C.  
; Jacobs, Kenneth  
; Hewick, Rodney M.  
; Genser, Thomas G.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/124,557  
; FILING DATE: 16-Apr-2002  
; CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann

; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 876-1170  
; TELEFAX: (617) 876-5851

; INFORMATION FOR SEQ ID NO: 46:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1320 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 46:

US-10-124-557-46

Query Match 99.7%; Score 5358; DB 13; Length 1320;  
Best Local Similarity 95.3%; Pred. No. 5, 6e-134;  
Matches 1006; Conservative 0; Mismatches 0; Indels 50; Gaps 1;  
QY 1 MAWKTLPIYLLLLLSVFIQVSSQELSCGRCFESFERGRCDCDAQCKKYDKCCPDYE 60  
Db 1 MAWKTLPIYLLLLLSVFIQVSSQELSCGRCFESFERGRCDCDAQCKKYDKCCPDYE 60  
QY 61 SPCA-----EVDKNK 70  
Db 61 SPCAEVHNPTSPSSSKKAPPPSGASQTIKSTTKRSPKPNKKTKKVIIESEITEVDKNK 120  
QY 71 KNRTKKPTPKPPVDEAGSLDNGDFKVTTPDTSTTQHNKYSTSPKIIITAKPINRPSL 130  
Db 121 KNRTKKPTPKPPVDEAGSLDNGDFKVTTPDTSTTQHNKYSTSPKIIITAKPINRPSL 180  
QY 131 PPNSDTSKETSITVANKETTVETKETTNNKQSTDGKEKTTSAKETQSIEKTSAKDLAPT 190  
Db 181 PPNSDTSKETSITVANKETTVETKETTNNKQSTDGKEKTTSAKETQSIEKTSAKDLAPT 240  
QY 191 SKVLAKPTPKAETTTKGPALITTPKEPTPTTPKEPASTTPEKPTPTTIKSAPITPKEPAPT 250  
Db 241 SKVLAKPTPKAETTTKGPALITTPKEPTPTTPKEPASTTPEKPTPTTIKSAPITPKEPAPT 300  
QY 251 TTKSAPITPKEPAPTTTKEPAPTTTPKEPAPTTTKSAPITTKSAPITPKEPAPTTPKKPA 310  
Db 301 TTKSAPITPKEPAPTTTKEPAPTTTPKEPAPTTTKSAPITTKSAPITPKEPAPTTPKKPA 360  
QY 311 PTTPKEPAPTTPEKPTPTTPKEPAPTTTPKEPAPTTAPKKAPAPTTPKEPAPTTTPK 370  
Db 361 PTTPKEPAPTTPEKPTPTTPKEPAPTTTPKEPAPTTAPKKAPAPTTPKEPAPTTTPK 420  
QY 371 BPAPTTTKPSPTTPKEPAPTTTKSAPTTTKSAPTTTKSAPTTTKSAPTTTKSAPTT 430  
Db 421 EPAPTTTKPSPTTPKEPAPTTTKSAPTTTKSAPTTTKSAPTTTKSAPTTTKSAPTT 480  
QY 431 PKEPAPTTPKKAPAPTTTPKEPAPTTTPKEPAPTTTKKAPATAPKEPAPTTPKETAPT 490  
Db 481 PKEPAPTTPKKAPAPTTTPKEPAPTTTPKEPAPTTTKKAPATAPKEPAPTTPKETAPT 540  
QY 491 TPPTPEKLAPTTPEKAPAPTTPELAPTTPEEPPTTPKEPAPTTPKAAANTPKEPAPTT 550  
Db 541 TPPTPEKLAPTTPEKAPAPTTPELAPTTPEEPPTTPKEPAPTTPKAAANTPKEPAPTT 600  
QY 551 PKEPAPTTKEPAPTTPKETAPTTTKGAPATTLKEPAPTTKPKAPAPKELAPTTTKEPTST 610  
Db 601 PKEPAPTTKEPAPTTPKETAPTTTKGAPATTLKEPAPTTKPKAPAPKELAPTTTKEPTST 660  
QY 611 TSDKAPAPTTPKGTAPTTTPKEPAPTTTPKGATPTTLKEPAPTTTPKKAPAPKELAP 670  
Db 661 TSDKAPAPTTPKGTAPTTTPKEPAPTTTPKGATPTTLKEPAPTTTPKKAPAPKELAP 720  
QY 671 TTTKGPTSTTSDKAPAPTTTPKETAPTTTPKEPAPTTTPKAPAPTTTETPTPTTSEVSTPTTK 730  
Db 721 TTTKGPTSTTSDKAPAPTTTPKETAPTTTPKEPAPTTTPKAPAPTTTETPTPTTSEVSTPTTK 780  
QY 731 EPTTIHKSPDESTPELSAETPKALENSKPEGVPTTKTPAATKPEMTTAKDKTTERDL 790  
Db 781 EPTTIHKSPDESTPELSAETPKALENSKPEGVPTTKTPAATKPEMTTAKDKTTERDL 840  
QY 791 RTTPTTAAAPKMTKETATTTKTESKITATTTQVTSSTTQDTPPKITTLKTTTLAPK 850  
Db 841 RTTPTTAAAPKMTKETATTTKTESKITATTTQVTSSTTQDTPPKITTLKTTTLAPK 900  
QY 851 VTTTKKTTITTEIMNKPEETAKPKORATNSKATTPKPKQKPTKAPKKPTSTKKPKMPVR 910  
Db 901 VTTTKKTTITTEIMNKPEETAKPKORATNSKATTPKPKQKPTKAPKKPTSTKKPKMPVR 960  
QY 911 KPKTTPTRKMTSTMPBLNPTSGRIASAMLOTTTRPNQTPNSKLVEVNPXSEDAGGABGET 970  
Db 961 KPKTTPTRKMTSTMPBLNPTSGRIASAMLOTTTRPNQTPNSKLVEVNPXSEDAGGABGET 1020



Qy 971 PHMLLRPHVFMPEVTPMDYLPVFNQGIINPMLS 1006  
 Db 1021 PHMLLRPHVFMPEVTPMDYLPVFNQGIINPMLS 1056

RESULT 7  
 US-10-124-557-60  
 ; Sequence 60, Application US/10124557  
 ; Publication No. US20020137894A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Turner, Katherine  
 ; Clark, Stephen C.  
 ; Jacobs, Kenneth  
 ; Hewick, Rodney M.  
 ; Gesner, Thomas G.  
 ;  
 ; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
 ; NUMBER OF SEQUENCES: 143  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Genetics Institute, Inc.  
 ; STREET: 87 CambridgePark Drive  
 ; CITY: Cambridge  
 ; STATE: Massachusetts  
 ; COUNTRY: U.S.A.  
 ; ZIP: 02140  
 ;  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ;  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/10/124,557  
 ; FILING DATE: 16-Apr-2002  
 ; CLASSIFICATION: <Unknown>  
 ;  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/643,502  
 ; FILING DATE: 18-JAN-1991  
 ; APPLICATION NUMBER: US 07/546,114  
 ; FILING DATE: 29-JUN-1990  
 ; APPLICATION NUMBER: US 07/457,196  
 ; FILING DATE: 29-DEC-1989  
 ; APPLICATION NUMBER: US 07/390,901  
 ; FILING DATE: 08-AUG-1989  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Cserr, Luann  
 ; REGISTRATION NUMBER: 31,822  
 ; REFERENCE/DOCKET NUMBER: GI 5190  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (617)876-1170  
 ; TELEFAX: (617)876-5851  
 ;  
 ; INFORMATION FOR SEQ ID NO: 60:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1320 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ;  
 ; MOLECULE TYPE: protein  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 60:  
 US-10-124-557-60

Query Match 99.7%; Score 5358; DB 13; Length 1320;  
 Best Local Similarity 95.3%; Pred. No. 5.6e-134;  
 Matches 1006; Conservative 0; Mismatches 0; Indels 50; Gaps 1;

Qy 1 MAWKTLPIYLLLLLSVFVIQQVSSQELSCGRCFESFERGRCDCDAQCKKYDKCCPYE 60  
 Db 1 MAWKTLPIYLLLLLSVFVIQQVSSQELSCGRCFESFERGRCDCDAQCKKYDKCCPYE 60

Qy 61 SPCA-----EVKDKK 70  
 Db 61 SPCA-----EVKDKK 70

Qy 71 KNRTKKKTPKPPVVDKAGSLDNGDFKVTTPDSTTQHNVKYSTSPKITTAKPINPRL 130  
 Db 121 KNRTKKKTPKPPVVDKAGSLDNGDFKVTTPDSTTQHNVKYSTSPKITTAKPINPRL 180

Qy 131 PPNSDTSKETSLTVNKEETTIVETKTTTTNNKQISTDGEKTTISAKETOSIEKTSAKDLAPT 190  
 Db 181 PPNSDTSKETSLTVNKEETTIVETKTTTTNNKQISTDGEKTTISAKETOSIEKTSAKDLAPT 240

Qy 191 SKVLAKPTPKAETTTKGPALTTTKEPTPTTKEPASTTTPKEPTTTTISAKETTTTKEPAPT 250  
 Db 241 SKVLAKPTPKAETTTKGPALTTTKEPTPTTKEPASTTTPKEPTTTTISAKETTTTKEPAPT 300

Qy 251 TTKSAPTTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPT 310  
 Db 301 TTKSAPTTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPT 360

Qy 311 PTTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPT 370  
 Db 361 PTTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPT 420

Qy 371 EPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPT 430  
 Db 421 EPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPT 480

Qy 431 PKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPT 490  
 Db 481 PKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPT 540

Qy 491 TPTTPEKLAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPT 550  
 Db 541 TPTTPEKLAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPT 600

Qy 551 PKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPT 610  
 Db 601 PKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPT 660

Qy 611 TSDKPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPT 670  
 Db 661 TSDKPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPT 720

Qy 671 TTTKGPTSTSDKPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPT 730  
 Db 721 TTTKGPTSTSDKPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPT 780

Qy 731 EPTTIHKSPDESTPELSAETPKALENSPKESGVPTTKTAAATKPEMTTAKDKTTTERDL 790  
 Db 781 EPTTIHKSPDESTPELSAETPKALENSPKESGVPTTKTAAATKPEMTTAKDKTTTERDL 840

Qy 791 RTTPTETTTAAPKMTKETATTTTEKTTESKITATTTQVTSITTTQDTTTPFKITTLKTTTLAPK 850  
 Db 841 RTTPTETTTAAPKMTKETATTTTEKTTESKITATTTQVTSITTTQDTTTPFKITTLKTTTLAPK 900

Qy 851 VTTTKTITTTTBMNKPETAKPKORATNSKATTPKQKPTKAPKPTSTKPKTMRVR 910  
 Db 901 VTTTKTITTTTBMNKPETAKPKORATNSKATTPKQKPTKAPKPTSTKPKTMRVR 960

Qy 911 KPKTTPTPKMTSTMPELNPTSRIAEAMLOTTTRNQTNSKLVVNPKSEBAGAGET 970  
 Db 961 KPKTTPTPKMTSTMPELNPTSRIAEAMLOTTTRNQTNSKLVVNPKSEBAGAGET 1020

Qy 971 PHMLLRPHVFMPEVTPMDYLPVFNQGIINPMLS 1006  
 Db 1021 PHMLLRPHVFMPEVTPMDYLPVFNQGIINPMLS 1056

RESULT 8  
 US-10-124-557-52  
 ; Sequence 52, Application US/10124557  
 ; Publication No. US20020137894A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Turner, Katherine  
 ; Clark, Stephen C.  
 ; Jacobs, Kenneth  
 ; Hewick, Rodney M.  
 ; Gesner, Thomas G.  
 ;  
 ; TITLE OF INVENTION: Megakaryocyte Stimulating Factors





```
;
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1361 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 40:
US-10-124-557-40

Query Match          99.5%; Score 5343.9; DB 13; Length 1361;
Best Local Similarity 91.7%; Pred. No. 1.4e-133;
Matches 1006; Conservative 0; Mismatches 0; Indels 91; Gaps 2;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25
Db 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGYSRDAICNCDYNCQHYMECCPDF 60
QY 26 -----ELSCGRCFFSFERGRCDCDAQCKYDKCCPDYESFCA----- 64
Db 61 KRVCIAELSCGRCFFSFERGRCDCDAQCKYDKCCPDYESFCAEVHNFTSPSSKKAP 120
QY 65 -----EVKONKKNRTKKKPKPKPPVVDVDEAG 89
Db 121 PPSGASQTIKSTKRSPPKPKKKTKVIESEBIETEVKONKKNRTKKKPKPKPPVVDVDEAG 180
QY 90 SGLDNGDFKVTTPDTSTQHNKSVSTSPKITTAKPINRPSLPNSDTSKETSLSVANKET 149
Db 181 SGLDNGDFKVTTPDTSTQHNKSVSTSPKITTAKPINRPSLPNSDTSKETSLSVANKET 240
QY 150 VETKETTTNKQSTDKGKTTSAKETQSTKTSKADLAPTSKVLAKPTPKAETTTKGPA 209
Db 241 VETKETTTNKQSTDKGKTTSAKETQSTKTSKADLAPTSKVLAKPTPKAETTTKGPA 300
QY 210 LTTPKKPTPTPKKPASTTKPKKPTTTIKSAPTTPKKPAPTTTKSAPTTPKKPAPTTKE 269
Db 301 LTTPKKPTPTPKKPASTTKPKKPTTTIKSAPTTPKKPAPTTTKSAPTTPKKPAPTTKE 360
QY 270 PAPTPPKKPAPTTTKPAPTTTKSAPTTPKKPAPTTTKPAPTTTKPAPTTTKPEPTPTT 329
Db 361 PAPTPPKKPAPTTTKPAPTTTKSAPTTPKKPAPTTTKPAPTTTKPAPTTTKPEPTPTT 420
QY 330 PKKPAPTTTKPAPTTTKPAPTTTKPAPTTTKPAPTTTKPAPTTTKPSPTTKKPA 389
Db 421 PKKPAPTTTKPAPTTTKPAPTTTKPAPTTTKPAPTTTKPAPTTTKPSPTTKKPA 480
QY 390 PTTTKSAPTTTKKPAPTTTKSAPTTPKKPSPTTKKPAPTTTKPAPTTTKPAPTTKE 449
Db 481 PTTTKSAPTTTKKPAPTTTKSAPTTPKKPSPTTKKPAPTTTKPAPTTTKPAPTTKE 540
QY 450 PAPTPPKKPAPTTTKPAPTTTKPAPTTTKPAPTTTKPAPTTTKPAPTTTKPAPTT 509
Db 541 PAPTPPKKPAPTTTKPAPTTTKPAPTTTKPAPTTTKPAPTTTKPAPTTTKPAPTT 600
QY 510 TPELAPTTPEPTPTTPEKPAPTTKAAADNTPKKPAPTTKKPAPTTKKPAPTTKKPA 569
Db 601 TPELAPTTPEPTPTTPEKPAPTTKAAADNTPKKPAPTTKKPAPTTKKPAPTTKKPA 660
QY 570 TAPTPKGTATTLKPAPTTPKKPAKELAPTTTKETSTSDKPAPTTTPKGTATTPK 629
Db 661 TAPTPKGTATTLKPAPTTPKKPAKELAPTTTKETSTSDKPAPTTTPKGTATTPK 720
QY 630 EPAPTPKPAPTTPKGTATTLKPAPTTPKKPAKELAPTTTKGTSTSDKPAPTTP 689
Db 721 EPAPTPKPAPTTPKGTATTLKPAPTTPKKPAKELAPTTTKGTSTSDKPAPTTP 780
```

```
QY 690 KETAPTTPKEPAPTTPKKPAPTTTPETPTTSEVSTPTTTKEPTTIHKSPDESTPELSAE 749
Db 781 KETAPTTPKEPAPTTPKKPAPTTTPETPTTSEVSTPTTTKEPTTIHKSPDESTPELSAE 840
QY 750 PTPKALENSPKFPGVPTTKTAAATKPEMTTAKDKTTERDLATTPETTTAAAPKMTKETAT 809
Db 841 PTPKALENSPKFPGVPTTKTAAATKPEMTTAKDKTTERDLATTPETTTAAAPKMTKETAT 900
QY 810 TTEKTTESKITATTTQVSTTTQDTTPPKITTLKTTILAPKVTTKKTTITTEIMNKPEE 869
Db 901 TTEKTTESKITATTTQVSTTTQDTTPPKITTLKTTILAPKVTTKKTTITTEIMNKPEE 960
QY 870 TAKPKDRATNSKATTPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPK 929
Db 961 TAKPKDRATNSKATTPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPK 1020
QY 930 PTSRIAEMLQTTTRPNQTPNSKLVEVNPVNSKEDAGGAEGETPHMLLRPHVFMPEVTPDMD 989
Db 1021 PTSRIAEMLQTTTRPNQTPNSKLVEVNPVNSKEDAGGAEGETPHMLLRPHVFMPEVTPDMD 1080
QY 990 YLPRVFNQGIINPMLS 1006
Db 1081 YLPRVFNQGIINPMLS 1097

RESULT 11
US-10-124-557-104
; Sequence 104, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1140 amino acids
```

```

;
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 104:
US-10-124-557-104

Query Match      99.4%; Score 5339.6; DB 13; Length 1140;
Best Local Similarity 88.2%; Pred. No. 1.5e-133;
Matches 1006; Conservative 0; Mismatches 0; Indels 134; Gaps 2;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25
Db 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 60
QY 26 -----ELSCGRCFCFESFERGECDDAQCKYDKCCPDYSEFCAE----- 65
Db 61 KRVCTAELSCGRCFCFESFERGECDDAQCKYDKCCPDYSEFCAEVHNFTSPSSKKAP 120
QY 66 ----- 65
Db 121 PPSGASQTIKSTTKRGPKNKKTKKVIESEBITEHVSSENQESSSSSSSSSTI 180
QY 66 -----VKONKQRTKKKPTPKPPVVDVDEAGSGLDNGDFKVTTPDTST 106
Db 181 KIKSSKNSAANRELOKKLVKDNKNKRTKKKPTPKPPVVDVDEAGSGLDNGDFKVTTPDTST 240
QY 107 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVNKETTVEKETTNNKQSTDG 166
Db 241 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVNKETTVEKETTNNKQSTDG 300
QY 167 KEKTTSAKETQSIEKTSKDLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTTPKEPAS 226
Db 301 KEKTTSAKETQSIEKTSKDLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTTPKEPAS 360
QY 227 TTPKEPTPTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKEPAPTTTKEP 286
Db 361 TTPKEPTPTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKEPAPTTTKEP 420
QY 287 APTTTKSAPTTPKEAPTTPKPAPTTPKPEAPTTTKEPTTTPKEAPTTTKEPAPTTPK 346
Db 421 APTTTKSAPTTPKEAPTTPKPAPTTPKPEAPTTTKEPTTTPKEAPTTTKEPAPTTPK 480
QY 347 EPAPTAPKAPAPTTPKPEAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 406
Db 481 EPAPTAPKAPAPTTPKPEAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 540
QY 407 TTKSAPTTPKESPTTKEPAPTTKPEAPTTPKPAPTTPKPEAPTTTKEPAPTTTKEP 466
Db 541 TTKSAPTTPKESPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 600
QY 467 APTAPKEAPTTPKETAPTTPKLTPPTPEKLAFTTTPKEAPTTTKEPAPTTTKEPAPTT 526
Db 601 APTAPKEAPTTPKETAPTTPKLTPPTPEKLAFTTTPKEAPTTTKEPAPTTTKEPAPTT 660
QY 527 PEEAPTTTAKAAPTTPKEAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 586
Db 661 PEEAPTTTAKAAPTTPKEAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 720
QY 587 APTTPEKAPKELAPTTTKEPTSTSDKAPPTPKGTAPTTPKPEAPTTTKEPAPTTTKEP 646
Db 721 APTTPEKAPKELAPTTTKEPTSTSDKAPPTPKGTAPTTPKPEAPTTTKEPAPTTTKEP 780
QY 647 TAPTTLKEPAPTTPKKAPKELAPTTTKEPTSTSDKAPPTPKGTAPTTPKPEAPTTTKEP 706
Db 781 TAPTTLKEPAPTTPKKAPKELAPTTTKEPTSTSDKAPPTPKGTAPTTPKPEAPTTTKEP 840
QY 707 KPAPTTTPEPTTSEVSTPTTKEPTTIHKSPEDESTPELSAEPPTPKALENSPKRGVPT 766
Db 841 KPAPTTTPEPTTSEVSTPTTKEPTTIHKSPEDESTPELSAEPPTPKALENSPKRGVPT 900
QY 767 TKTPAATKPEMTTAKDKITRDLRTTPEPTTAAAPKMTKETATTTTEKTESKITATTTOV 826
Db 901 TKTPAATKPEMTTAKDKITRDLRTTPEPTTAAAPKMTKETATTTTEKTESKITATTTOV 960

```

```

QY 827 TSTTTQDTPPKITTLTKTTLAPKVTTKTITTEIMNKPEETAKPKDRATNSKATTPK 886
Db 961 TSTTTQDTPPKITTLTKTTLAPKVTTKTITTEIMNKPEETAKPKDRATNSKATTPK 1020
QY 887 PQKPTKAPKPTSTKPKMTKPRVRKPKTTPTPPKMTSTMPELNPTSHIAEAMLOTTTRPN 946
Db 1021 PQKPTKAPKPTSTKPKMTKPRVRKPKTTPTPPKMTSTMPELNPTSHIAEAMLOTTTRPN 1080
QY 947 QTNSKLVEVNPKSESDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPVNPQGIINPMLS 1006
Db 1081 QTNSKLVEVNPKSESDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPVNPQGIINPMLS 1140

RESULT 12
US-09-802-207-30
; Sequence 30, Application US/09802207
; Publication No. US20020086824A1
; GENERAL INFORMATION:
; APPLICANT: Warman, Matthew
; APPLICANT: Carpten, John
; APPLICANT: Trent, Jeffrey
; APPLICANT: Marcelino, Jose
; TITLE OF INVENTION: Novel Methods and Reagents for the Treatment of Osteoarthritis
; FILE REFERENCE: Case-06212
; CURRENT APPLICATION NUMBER: US/09/802,207
; CURRENT FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 09/619,175
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 60/145,328
; PRIOR FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 30
; LENGTH: 1404
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-802-207-30

```

```

Query Match      99.4%; Score 5339.6; DB 9; Length 1404;
Best Local Similarity 88.2%; Pred. No. 1.8e-133;
Matches 1006; Conservative 0; Mismatches 0; Indels 134; Gaps 2;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25
Db 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 60
QY 26 -----ELSCGRCFCFESFERGECDDAQCKYDKCCPDYSEFCAE----- 65
Db 61 KRVCTAELSCGRCFCFESFERGECDDAQCKYDKCCPDYSEFCAEVHNFTSPSSKKAP 120
QY 66 ----- 65
Db 121 PPSGASQTIKSTTKRSPKPPNKKTKKVIIESEITEHVSSENQESSSSSSSSSTI 180
QY 66 -----VKONKQRTKKKPTPKPPVVDVDEAGSGLDNGDFKVTTPDTST 106
Db 181 KIKSSKNSAANRELOKKLVKDNKNKRTKKKPTPKPPVVDVDEAGSGLDNGDFKVTTPDTST 240
QY 107 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVNKETTVEKETTNNKQSTDG 166
Db 241 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVNKETTVEKETTNNKQSTDG 300
QY 167 KEKTTSAKETQSIEKTSKDLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTTPKEPAS 226
Db 301 KEKTTSAKETQSIEKTSKDLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTTPKEPAS 360
QY 227 TTPKEPTPTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKEPAPTTTKEP 286
Db 361 TTPKEPTPTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKEPAPTTTKEP 420
QY 287 APTTTKSAPTTPKEAPTTPKPAPTTPKPEAPTTTKEPTTTPKEAPTTTKEPAPTTPK 346

```



Db 1081 QTPNSKLVNPKSEDAGGAGETPHMLLRPHVFMPEVTPDMOYLPRVFNQGIINPMLS 1140  
 |||||  
 RESULT 14  
 US-10-124-557-2  
 ; Sequence 2, Application US/10124557  
 ; Publication No. US20020137894A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Turner, Katherine  
 ; Clark, Stephen C.  
 ; Jacobs, Kenneth  
 ; Hewick, Rodney M.  
 ; Gesner, Thomas G.  
 ; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
 ; NUMBER OF SEQUENCES: 143  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Genetics Institute, Inc.  
 ; STREET: 87 Cambridgepark Drive  
 ; CITY: Cambridge  
 ; STATE: Massachusetts  
 ; COUNTRY: U.S.A.  
 ; ZIP: 02140  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/10/124,557  
 ; Filing DATE: 16-Apr-2002  
 ; CLASSIFICATION: <Unknown>  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/643,502  
 ; Filing DATE: 18-JAN-1991  
 ; APPLICATION NUMBER: US 07/546,114  
 ; Filing DATE: 29-JUN-1990  
 ; APPLICATION NUMBER: US 07/457,196  
 ; Filing DATE: 29-DEC-1989  
 ; APPLICATION NUMBER: US 07/390,901  
 ; Filing DATE: 08-AUG-1989  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Cserr, Luann  
 ; REGISTRATION NUMBER: 31,822  
 ; REFERENCE/DOCKET NUMBER: GI 5190  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (617) 876-1170  
 ; TELEFAX: (617) 876-5851  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1404 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
 US-10-124-557-2  
 Query Match 99.4%; Score 5339.6; DB 13; Length 1404;  
 Best Local Similarity 88.2%; Pred. No. 1.8e-133;  
 Matches 1006; Conservative 0; Mismatches 0; Indels 134; Gaps 2;  
 Qy 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25  
 Db 1 MAWKTLPIYLLLLSVFVIQVSSQDLSACGRCGYSRDATCNCDYNCQHYMECCPDF 60  
 Qy 26 -----ELCKGRCPESFERGECDCDAQCKYDKCCPDYEFCAE----- 65  
 Db 61 KRVCTAELSCRCRCPESFERGECDCDAQCKYDKCCPDYEFCAE VHNPTSPSSKKAP 120  
 Qy 66 ----- 65  
 Db 121 PPSGASQIKSTTKSPKPPNKKTKKVI ESEITEHSVSENQESSSSSSSSSSSTIW 180

Qy 66 -----VKONKNRTKKKPTPKPPVVD EAGSLONGDFKVTTPDTST 106  
 Db 181 KIKSSKNSAAMRELQKKLVKDNKNRTKKKPTPKPPVVD EAGSLONGDFKVTTPDTST 240  
 Qy 107 TOHNKVSTSPKILTTAKDINPRPSLPNSDTSKETSLTVNKETTVETKETTNTKQTSDDG 166  
 Db 241 TOHNKVSTSPKILTTAKDINPRPSLPNSDTSKETSLTVNKETTVETKETTNTKQTSDDG 300  
 Qy 167 KEKTTSAKETQSIEKTSADLAPTSKVLAKPTPKAETTTKGPA LTTTKEPTPTTKGPAS 226  
 Db 301 KEKTTSAKETQSIEKTSADLAPTSKVLAKPTPKAETTTKGPA LTTTKEPTPTTKGPAS 360  
 Qy 227 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKGPA LTTTKEPTPTTKGP 286  
 Db 361 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKGPA LTTTKEPTPTTKGP 420  
 Qy 287 APTTTKSAPTTPKEPAPTTTKGPA LTTTKEPTPTTKGPA LTTTKEPTPTTKGPAPTTPK 346  
 Db 421 APTTTKSAPTTPKEPAPTTTKGPA LTTTKEPTPTTKGPA LTTTKEPTPTTKGPAPTTPK 480  
 Qy 347 EPAPTAPKKEPAPTTTPKEPAPTTTPKEPAPTTTKGPA LTTTKEPTPTTKGPA LTTTKEPTPTTKGPA 406  
 Db 481 EPAPTAPKKEPAPTTTPKEPAPTTTPKEPAPTTTKGPA LTTTKEPTPTTKGPA LTTTKEPTPTTKGPA 540  
 Qy 407 TTKSAPTTPKEPSPTTKGPA LTTTKEPAPTTTPKEPAPTTTPKEPAPTTTKGPA LTTTKEPTPTTKGPA 466  
 Db 541 TTKSAPTTPKEPSPTTKGPA LTTTKEPAPTTTPKEPAPTTTPKEPAPTTTKGPA LTTTKEPTPTTKGPA 600  
 Qy 467 APTAPKEPAPTTPKETAPTTPKKLTPTTPEKLA PTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPTPTT 526  
 Db 601 APTAPKEPAPTTPKETAPTTPKKLTPTTPEKLA PTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPTPTT 660  
 Qy 527 PEEPAPTTPKAAAPNTPKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPTPTT 586  
 Db 661 PEEPAPTTPKAAAPNTPKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPTPTT 720  
 Qy 587 APTTPKKPAPKELAPTTTKEPTSTSDKAPTTPKGTAPTTTPKEPAPTTTPKEPAPTTTPKEPTPTT 646  
 Db 721 APTTPKKPAPKELAPTTTKEPTSTSDKAPTTPKGTAPTTTPKEPAPTTTPKEPAPTTTPKEPTPTT 780  
 Qy 647 TAPTTLKBPAPTTPKKAPKELAPTTTKEPTSTSDKAPTTTPKEPAPTTTPKEPAPTTTPKEPTPTT 706  
 Db 781 TAPTTLKBPAPTTPKKAPKELAPTTTKEPTSTSDKAPTTTPKEPAPTTTPKEPAPTTTPKEPTPTT 840  
 Qy 707 KPAPTTTPPTTSEVSTPTTKEPTTIHKSPDESTPELSA EPTPKALENSPKEPGVPT 766  
 Db 841 KPAPTTTPPTTSEVSTPTTKEPTTIHKSPDESTPELSA EPTPKALENSPKEPGVPT 900  
 Qy 767 TKTPAATKPEMTTAKDKTTERDLTTTPETTTAAAPMTKEATTTTEKTTESKITATTQV 826  
 Db 901 TKTPAATKPEMTTAKDKTTERDLTTTPETTTAAAPMTKEATTTTEKTTESKITATTQV 960  
 Qy 827 TSTTTQDTTTPPKITLTKTTLAPKVTTTKITTTIMNKPEETAKPKDRATNSKATTPK 886  
 Db 961 TSTTTQDTTTPPKITLTKTTLAPKVTTTKITTTIMNKPEETAKPKDRATNSKATTPK 1020  
 Qy 887 POKPTKAPKXTSTKPKKIMPRVRKPTTTPTRKMTSTWPELNP TSRIAEMLOTTTRPN 946  
 Db 1021 POKPTKAPKXTSTKPKKIMPRVRKPTTTPTRKMTSTWPELNP TSRIAEMLOTTTRPN 1080  
 Qy 947 QTPNSKLVNPKSEDAGGAGETPHMLLRPHVFMPEVTPDMOYLPRVFNQGIINPMLS 1006  
 Db 1081 QTPNSKLVNPKSEDAGGAGETPHMLLRPHVFMPEVTPDMOYLPRVFNQGIINPMLS 1140  
 RESULT 15  
 US-10-124-557-62  
 ; Sequence 62, Application US/10124557  
 ; Publication No. US20020137894A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Turner, Katherine  
 ; Clark, Stephen C.  
 ; Jacobs, Kenneth



Hewick, Rodney M.  
Gesner, Thomas G.  
TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
NUMBER OF SEQUENCES: 143  
CORRESPONDENCE ADDRESS:  
ADDRESS: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA: US/10/124,557  
FILING DATE: 16-Apr-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Cseri, Luann  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 5190  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 876-1170  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 62:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1404 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 62:  
US-10-124-557-62  
Query Match 99.4%; Score 5339.6; DB 13; Length 1404;  
Best Local Similarity 88.2%; Pred. No. 1.8e-133;  
Matches 1006; Conservative 0; Mismatches 0; Indels 134; Gaps 2;  
QY 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25  
Db 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRGEGYSRDATCNCYNCQHYMECCPDF 60  
QY 26 -----ELSCKGRCFESFERGECDDAQCKKYDKCCPDYESFCAE----- 65  
Db 61 KRVCVTAELSCKGRCFESFERGECDDAQCKKYDKCCPDYESFCAEVHNFTSPSSKAP 120  
QY 66 ----- 65  
Db 121 PPSGASQTIKSTTKRSPKPPNKKTKKVIREEITEEHSVSENQESSSSSSSSSTIW 180  
QY 66 -----VKDNKNRTKKKPTKPPVVDDEAGSLDNGDFKVTTPDTST 106  
Db 181 KIKSSKNSAANRELQKKLVKDNKNRTKKKPTKPPVVDDEAGSLDNGDFKVTTPDTST 240  
QY 107 TQHNKVSTSPKITTAKPINRPSLPNPSDTSKETSITVKNKETTIVETKETTITNKQSTDG 166  
Db 241 TQHNKVSTSPKITTAKPINRPSLPNPSDTSKETSITVKNKETTIVETKETTITNKQSTDG 300  
QY 167 KEKTTSAKETQSIEKTSKDLAPTSKVLAKEPTPKAETTKGPAITTKPEPTPTPKBPAS 226  
Db 301 KEKTTSAKETQSIEKTSKDLAPTSKVLAKEPTPKAETTKGPAITTKPEPTPTPKBPAS 360

QY 227 TTPKEPTTTIKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTKP 286  
Db 361 TTPKEPTTTIKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTKP 420  
QY 287 APTTTKSAPTTTPKEPAPTTTPKAPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 346  
Db 421 APTTTKSAPTTTPKEPAPTTTPKAPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 480  
QY 347 EPAPAPKAPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 406  
Db 481 EPAPAPKAPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 540  
QY 407 TTKSAPTTTPKEPAPTTTPKEPAPTTTPKAPAPTTTPKAPAPTTTPKAPAPTTTPKAPAPTTTPKAP 466  
Db 541 TTKSAPTTTPKEPAPTTTPKEPAPTTTPKAPAPTTTPKAPAPTTTPKAPAPTTTPKAPAPTTTPKAP 600  
QY 467 APTAPKEPAPTTTPKETAPTTPKLTPPTPEKAPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTP 526  
Db 601 APTAPKEPAPTTTPKETAPTTPKLTPPTPEKAPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTP 660  
QY 527 PEEAPAPTTPKAAPNTPKAPAPTTTPKEPAPTTTPKEPAPTTTPKETAPTTTPKETAPTTTPKETAPTTTPKET 586  
Db 661 PEEAPAPTTPKAAPNTPKAPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKETAPTTTPKETAPTTTPKET 720  
QY 587 APTTPKAPAPKELAPTTTKETPTSTSDKAPAPTTTPKAPAPTTTPKAPAPTTTPKAPAPTTTPKAPAPTTTPK 646  
Db 721 APTTPKAPAPKELAPTTTKETPTSTSDKAPAPTTTPKAPAPTTTPKAPAPTTTPKAPAPTTTPKAPAPTTTPK 780  
QY 647 TAPTTIKAPAPTTTPKAPAPKELAPTTTKGPTSTSDKAPAPTTTPKAPAPTTTPKAPAPTTTPKAPAPTTTPK 706  
Db 781 TAPTTIKAPAPTTTPKAPAPKELAPTTTKGPTSTSDKAPAPTTTPKAPAPTTTPKAPAPTTTPKAPAPTTTPK 840  
QY 707 KAPAPTTTPETPPTTSEVSTPTTKETPTTHKSPDESTPELSAPTTPELSAPTTPELSAPTTPELSAPTTPELS 766  
Db 841 KAPAPTTTPETPPTTSEVSTPTTKETPTTHKSPDESTPELSAPTTPELSAPTTPELSAPTTPELSAPTTPELS 900  
QY 767 TKTPAATKPEMTTAKDKTTERDLRTTPETTTAAPKMTKETATTTTETTESKITATTITQV 826  
Db 901 TKTPAATKPEMTTAKDKTTERDLRTTPETTTAAPKMTKETATTTTETTESKITATTITQV 960  
QY 827 TSTTTQDTPPFKITTILKTTLAPKVTTTKKTIITTEIMNKPEETAKPKDRATNSKATTPK 886  
Db 961 TSTTTQDTPPFKITTILKTTLAPKVTTTKKTIITTEIMNKPEETAKPKDRATNSKATTPK 1020  
QY 887 POKPTKAPKPTSTKPKMPVRKPTTPPKMTSTMPELNPTSRIAEAMLTQTTTRPN 946  
Db 1021 POKPTKAPKPTSTKPKMPVRKPTTPPKMTSTMPELNPTSRIAEAMLTQTTTRPN 1080  
QY 947 QTPNSKLVEVNPKSSEDAGAGETPHMLLRPHVFMPEVTPDMOYLPRVPNQGIINPMLS 1006  
Db 1081 QTPNSKLVEVNPKSSEDAGAGETPHMLLRPHVFMPEVTPDMOYLPRVPNQGIINPMLS 1140

Search completed: October 13, 2004, 11:53:08  
Job time : 112.349 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 110.61 Seconds  
(without alignments)  
5233.063 Million cell updates/sec

Title: SEQ1-E  
Perfect score: 5373  
Sequence: 1 MAWKTLPIYLLLLSVFVIQ.....DMDYLPRVNPQGIINPMLS 1006

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.1

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Uniprot\_02:\*

1: uniprot\_sprot:\*

2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	5339.6	99.4	1404	2	Q92954	Q92954 homo sapien
2	5330.6	99.2	1404	2	Q9BX49	Q9BX49 homo sapien
3	2949.7	54.9	933	2	Q62M25	Q62M25 homo sapien
4	2949.7	54.9	933	2	BAD18580	BAD18580 h cDNA fl
5	2307.7	42.9	1054	2	Q9JN99	Q9JN99 mus musculus
6	1335.3	24.9	5179	1	MUC2_HUMAN	MUC2_HUMAN
7	1306.5	24.3	1225	2	Q9VR49	Q9VR49 homo sapien
8	1299.5	24.2	1761	2	Q7KTF6	Q7KTF6 drosophila
9	1299.5	24.2	1761	2	AAS64673	AAS64673 drosophila
10	1227.8	22.9	3150	2	Q7PMD5	Q7PMD5 anopheles g
11	1141.9	21.3	3409	2	Q6SSE6	Q6SSE6 chlamydomon
12	1141.9	21.3	3409	2	AAS07044	AAS07044 chlamydomon
13	1108.6	20.6	1664	1	SDP1_CLOTTM	SDP1_CLOTTM
14	1102.5	20.5	1349	2	Q8WWQ4	Q8WWQ4 homo sapien
15	1082.4	20.1	3432	2	Q8TR51	Q8TR51 drosophila
16	1082.4	20.1	3458	2	Q8TR52	Q8TR52 drosophila
17	1069.7	19.9	3889	2	Q6SSE8	Q6SSE8 chlamydomon
18	1069.7	19.9	3889	2	AAS07042	AAS07042 chlamydomon
19	1041.7	19.4	1795	2	Q76894	Q76894 drosophila
20	1038.2	19.3	1079	2	Q9N4S7	Q9N4S7 caenorhabdi
21	1023.4	19.0	972	2	Q7OKV7	Q7OKV7 anopheles g
22	1014.7	18.9	1607	2	Q8H6Q5	Q8H6Q5 phytophthor
23	1014.7	18.9	1607	2	AAP74661	AAP74661 phytophthor
24	1008.8	18.8	2284	2	Q9VPG1	Q9VPG1 drosophila
25	1002.8	18.7	1274	2	Q20007	Q20007 caenorhabdi
26	998.5	18.6	9234	2	Q7KTP5	Q7KTP5 drosophila
27	998.5	18.6	9234	2	AAN10531	AAN10531 drosophila
28	990	18.4	1489	2	Q9F449	Q9F449 phytophthor
29	977	18.2	23015	2	Q8IQ18	Q8IQ18 drosophila
30	977	18.2	23015	2	AAN10358	AAN10358 drosophila
31	974.3	18.1	34350	2	Q8WZ42	Q8WZ42 homo sapien

## RESULT 1

ID	Q92954	PRELIMINARY;	PRT;	1404	AA.
AC	Q92954;				
DT	01-FEB-1997	(TrEMBLrel. 02, Created)			
DT	01-FEB-1997	(TrEMBLrel. 02, Last sequence update)			
DT	01-MAR-2004	(TrEMBLrel. 26, Last annotation update)			
DE	Megakaryocyte stimulating factor.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]_TaxID=9606;				
RP	SEQUENCE FROM N.A.				
RA	Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,				
RA	Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,				
RA	Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,				
RA	Bhatia S., Kriz R., Hewick R., Clark S.C.;				
RT	"Purification, Biochemical Characterization, and Cloning of a Novel				
RT	Megakaryocyte Stimulating Factor that has Megakaryocyte Colony				
RT	Stimulating Activity.";				
RL	Blood 78:279-279(1991).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RA	Merberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P.,				
RA	Fitzgerald M., Scaltreto J., Kelleher K., Preissner K., Kriz R.,				
RA	Jacobs K., Turner K.;				
RT	"A Comparison of Vitronectin and Megakaryocyte Stimulating Factor.";				
RL	Mosher D.F. (eds.);				
RL	BIOLOGY OF VITRONECTINS AND THEIR RECEPTORS., pp.45-52, Elsevier				
RL	Science Publishers B.V. (1993).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RA	Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,				
RA	Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,				
RA	Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,				
RA	Bhatia S., Kriz R., Hewick R., Clark S.C.;				
RL	Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; U70136; AAB09089.1; -.				
DR	HSSP; P04004; 10C0.				
DR	Genew; HGNC:9364; PRG4.				
DR	GO; GO:0008283; P:cell proliferation; TAS.				
DR	InterPro; IPR000585; Hemopexin.				
DR	InterPro; IPR001212; Somatomedin_B.				
DR	Pfam; PF00045; Hemopexin; 2.				
DR	PRINTS; PR00022; Somatomedin_B; 2.				
DR	SMART; SM00120; HX; 2.				
DR	SMART; SM00201; SO; 2.				
DR	PROSITE; PS00024; HEMOPEXIN; UNKNOWN 1.				
DR	PROSITE; PS00524; SOMATOMEDIN_B; 2.				
DR	SEQUENCE 1404 AA; 151090 MW; AABD7AD19B35F4F6 CRC64;				

## ALIGNMENTS

32	973.9	18.1	7962	2	Q10465	Q10465 homo sapien
33	969.9	18.1	2187	2	P70670	P70670 mus musculus
34	966.3	18.0	10578	2	Q8ISF5	Q8ISF5 caenorhabdi
35	964.3	17.9	18519	2	Q8ISF6	Q8ISF6 caenorhabdi
36	964.3	17.9	18534	2	Q8ISF7	Q8ISF7 caenorhabdi
37	960.1	17.9	1458	2	Q757N5	Q757N5 ashbya goss
38	960.1	17.9	1458	2	AAS52662	AAS52662 ashbya go
39	956.9	17.8	1121	2	Q7Z884	Q7Z884 candida alb
40	955.4	17.8	5703	1	MUSB_HUMAN	MUSB_HUMAN
41	955.1	17.8	1480	2	Q9LIE8	Q9LIE8 arabisidopsis
42	948.6	17.7	2112	2	Q9VEL9	Q9VEL9 drosophila
43	946.8	17.6	926	2	Q9VVG2	Q9VVG2 drosophila
44	941.6	17.5	1720	2	Q8I486	Q8I486 plasmodium
45	922.2	17.2	2042	2	Q767L8	Q767L8 sus scrofa

```
Query Match          99.4%; Score 5339.6; DB 2; Length 1404;
Best Local Similarity 88.2%; Pred. No. 28-97;
Matches 1006; Conservative 0; Mismatches 0; Indels 134; Gaps 2;

QY 1 MAWKTLPIYLLLLSVFVIQQVSSQ----- 25
DB 1 MAWKTLPIYLLLLSVFVIQQVSSQDLSSCAGRCGYSRDATCNCYNCQHMECCPDF 60

QY 26 -----ELSKGRCFESFERGECDCDAQCKYDKCCPDYESFCAB----- 65
DB 61 KRVCTAELSCGRCFESFERGECDCDAQCKYDKCCPDYESFCABVHNFTSPSSKKAP 120

QY 66 ----- 65
DB 121 PPSGASQTIKSTTKRSPKPNKKTKVIESEBITTEHSVSNQESSSSSSSSSTIR 180

QY 66 -----VKONKNRTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDTST 106
DB 181 KIKSSKNSAANRELQKKLVKONKNRTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDTST 240

QY 107 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKETTVETKETTNNKQSTDG 166
DB 241 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKETTVETKETTNNKQSTDG 300

QY 167 KEKTTSAKETQSLEKTSADKLAFTSKVLAKPTPKAETTTKGPALTTTPKESPTTPPKEPAS 226
DB 301 KEKTTSAKETQSLEKTSADKLAFTSKVLAKPTPKAETTTKGPALTTTPKESPTTPPKEPAS 360

QY 227 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKEP 286
DB 361 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKEP 420

QY 287 APTTTKSAPTTPKEPAPTTTPKAPPTPKAPPTPKAPPTTPKEPAPTTTPKEPAPTTTPK 346
DB 421 APTTTKSAPTTPKEPAPTTTPKAPPTPKAPPTPKAPPTTPKEPAPTTTPKEPAPTTTPK 480

QY 347 EPAPTAPKAPPTTPKEPAPTTTPKEPAPTTTKPSPTTPKEPAPTTTKSAPTTTKAPPTTK 406
DB 481 EPAPTAPKAPPTTPKEPAPTTTPKEPAPTTTKPSPTTPKEPAPTTTKSAPTTTKAPPTTK 540

QY 407 TTKSAPTTPKEPSPTTKAPPTTPKEPAPTTTPKXPAPTTTPKEPAPTTTPKEPAPTTTKKP 466
DB 541 TTKSAPTTPKEPSPTTKAPPTTPKEPAPTTTPKXPAPTTTPKEPAPTTTPKEPAPTTTKKP 600

QY 467 APTAPKEPAPTTTPKEPAPTTPKLTPPTPKLAPPTTPKEPAPTTPELAPTTPEEBPTPT 526
DB 601 APTAPKEPAPTTTPKEPAPTTPKLTPPTPKLAPPTTPKEPAPTTPELAPTTPEEBPTPT 660

QY 527 PEPAPPTPKAAAPNTPKAPPTTPKEPAPTTTPKEPAPTTPKETAPTTPKGTAPTTLKEP 586
DB 661 PEPAPPTPKAAAPNTPKAPPTTPKEPAPTTTPKEPAPTTPKETAPTTPKGTAPTTLKEP 720

QY 587 APTTPKAPKAPKELAPTTTKETSTSDKAPPTTPKGTAPTTTPKEPAPTTTPKEPAPTTPKG 646
DB 721 APTTPKAPKAPKELAPTTTKETSTSDKAPPTTPKGTAPTTTPKEPAPTTTPKEPAPTTPKG 780

QY 647 TAPPTLKAPPTTPKAPKELAPTTTKGTSTSDKAPPTTPKAPTTTPKEPAPTTTPKEPAPTTPK 706
DB 781 TAPPTLKAPPTTPKAPKELAPTTTKGTSTSDKAPPTTPKAPTTTPKEPAPTTTPKEPAPTTPK 840

QY 707 KPAPPTTPPETPTTSVSTPTTTKEPTTIHKSPPDESTPELSAEPPTPKALENSPKPGVPT 766
DB 841 KPAPPTTPPETPTTSVSTPTTTKEPTTIHKSPPDESTPELSAEPPTPKALENSPKPGVPT 900

QY 767 TKTPAATKPBMTTTAKDKTERDLRTPTTTTAAPKMKETATTTKETSKITATTQV 826
DB 901 TKTPAATKPBMTTTAKDKTERDLRTPTTTTAAPKMKETATTTKETSKITATTQV 960

QY 827 TSTTTQDTPPKKITTLKTTILAPKVTTKKTIITTEIMNKPEETAKPKDRATNSKATTPK 886
DB 961 TSTTTQDTPPKKITTLKTTILAPKVTTKKTIITTEIMNKPEETAKPKDRATNSKATTPK 1020

QY 887 PQKFTKAPKKTSTKPKTMPRVKPKTTTPTRKMTSTMPNLNPTSRIAEAMLQTTTRPN 946
```

```
DB 1021 PQKPIKPKPTISTKKPTMPRVKPKTTTPTRKMTSTMPNLNPTSRIAEAMLQTTTRN 1080
QY 947 QTPNSKLVEVNPDKSDAGAGETPHMLLRPHVFPVEVTPDMDYLPVFNQGIINPMLS 1006
DB 1081 QTPNSKLVEVNPDKSDAGAGETPHMLLRPHVFPVEVTPDMDYLPVFNQGIINPMLS 1140

RESULT 2
Q9BX49 PRELIMINARY; PRT; 1404 AA.
AC Q9BX49;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE BG174L6.2 (MSF: megakaryocyte stimulating factor ).
GN Name=BG174L6.2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wray P.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL133553; CAC36090.1; -.
DR HSSP; P04004; LOCO.
DR InterPro; IPR000585; Hemopexin.
DR InterPro; IPR001212; Somatomedin_B.
DR Pfam; PF00045; Hemopexin; 2.
DR Pfam; PF01033; Somatomedin B; 2.
DR PRINTS; PR00022; SOMATOMEDINB.
DR SMART; SM00120; HX; 2.
DR SMART; SM00201; SO; 2.
DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR PROSITE; PS00524; SOMATOMEDIN B; 2.
DR PROSITE; PS00524; SOMATOMEDIN B; 2.
SQ SEQUENCE 1404 AA; 151076 MW; 782A11746B3FDEE5 CRC64;

Query Match          99.2%; Score 5330.6; DB 2; Length 1404;
Best Local Similarity 88.1%; Pred. No. 3e-97;
Matches 1004; Conservative 0; Mismatches 0; Indels 134; Gaps 2;

QY 1 MAWKTLPIYLLLLSVFVIQQVSSQ----- 25
DB 1 MAWKTLPIYLLLLSVFVIQQVSSQDLSSCAGRCGYSRDATCNCYNCQHMECCPDF 60

QY 26 -----ELSKGRCFESFERGECDCDAQCKYDKCCPDYESFCAB----- 65
DB 61 KRVCTAELSCGRCFESFERGECDCDAQCKYDKCCPDYESFCABVHNFTSPSSKKAP 120

QY 66 ----- 65
DB 121 PPSGASQTIKSTTKRSPKPNKKTKVIESEBITTEHSVSNQESSSSSSSSSSSTIR 180

QY 66 -----VKONKNRTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDTST 106
DB 181 KIKSSKNSAANRELQKKLVKONKNRTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDTST 240

QY 107 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKETTVETKETTNNKQSTDG 166
DB 241 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKETTVETKETTNNKQSTDG 300

QY 167 KEKTTSAKETQSLEKTSADKLAFTSKVLAKPTPKAETTTKGPALTTTPKESPTTPPKEPAS 226
DB 301 KEKTTSAKETQSLEKTSADKLAFTSKVLAKPTPKAETTTKGPALTTTPKESPTTPPKEPAS 360

QY 227 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKEP 286
DB 361 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKEP 420

QY 287 APTTTKSAPTTPKEPAPTTTPKAPPTPKAPPTPKAPPTTPKEPAPTTTPKEPAPTTTPK 346
DB 421 APTTTKSAPTTPKEPAPTTTPKAPPTPKAPPTPKAPPTTPKEPAPTTTPKEPAPTTTPK 480
```

QY 347 EPAFTAPKPAFTTPKEPAFTTPKEPAFTTTKESPTTPKEPAFTTTKSAFTTTKEPAFT 406  
 DB 481 EPAFTAPKPAFTTPKEPAFTTPKEPAFTTTKESPTTPKEPAFTTTKSAFTTTKEPAFT 540  
 QY 407 TTKSAFTTPKEPSSTTTKEPAFTTPKEPAFTTPKKAFTTPKKAFTTPKKAFTTTTKP 466  
 DB 541 TTKSAFTTPKEPSSTTTKEPAFTTPKEPAFTTPKKAFTTPKKAFTTPKKAFTTTTKP 600  
 QY 467 APTAPKEPAFTTPKETAPTTPKLTTPKEPAFTTPKEPAFTTPKEPAFTTPPELAPTTPPEPTPT 526  
 DB 601 APTTPKEPAFTTPKETAPTTPKLTTPKEPAFTTPKEPAFTTPKEPAFTTPPELAPTTPPEPTPT 660  
 QY 527 PEEPAFTTPKAAAFNTTPKEPAFTTPKEPAFTTPKEPAFTTPKETAPTTPKGTAPTTLKEP 586  
 DB 661 PEEPAFTTPKAAAFNTTPKEPAFTTPKEPAFTTPKEPAFTTPKETAPTTPKGTAPTTLKEP 720  
 QY 587 APTTPKPAKELAPTTPKEPTSTSDKPAFTTPKGTAPTTPKPAFTTPKPAFTTPKPAFTTPKG 646  
 DB 721 APTTPKPAKELAPTTPKEPTSTSDKPAFTTPKGTAPTTPKPAFTTPKPAFTTPKPAFTTPKG 780  
 QY 647 TAPTTLKEPAFTTPKPAKELAPTTPKGTSTSDKPAFTTPKGTAPTTPKPAFTTPKPAFTTPK 706  
 DB 781 TAPTTLKEPAFTTPKPAKELAPTTPKGTSTSDKPAFTTPKGTAPTTPKPAFTTPKPAFTTPK 840  
 QY 707 KPAPTTPETPTTSEVSTPTTKEPTTIHKSPDESPELSAFTTPKALENSKPEPGVPT 766  
 DB 841 KPAPTTPETPTTSEVSTPTTKEPTTIHKSPDESPELSAFTTPKALENSKPEPGVPT 900  
 QY 767 TKTPAATKEMTTAKDKTTERDLRTPTTETTTAAPKMTKETATTTTEKTSKITATTQV 826  
 DB 901 TKTPAATKEMTTAKDKTTERDLRTPTTETTTAAPKMTKETATTTTEKTSKITATTQV 960  
 QY 827 TSTTTQDTTPFKLTTLTKTTLAPKVTITTKTITTTIMNKPSTAKPKORATNSKATTPK 886  
 DB 961 TSTTTQDTTPFKLTTLTKTTLAPKVTITTKTITTTIMNKPSTAKPKORATNSKATTPK 1020  
 QY 887 POKPTKAPKPTSTKPKTTPVRKPTTPTRKMTSTWPELNPTSGRIAEAMLOTTTRPN 946  
 DB 1021 POKPTKAPKPTSTKPKTTPVRKPTTPTRKMTSTWPELNPTSGRIAEAMLOTTTRPN 1080  
 QY 947 QTPNSKLVEVPKSDAGAGSTPHMLRPHVFNPEVTPDMDYLPRVFNQGIINPMLS 1006  
 DB 1081 QTPNSKLVEVPKSDAGAGSTPHMLRPHVFNPEVTPDMDYLPRVFNQGIINPMLS 1140

RESULT 3  
 Q6ZM25 PRELIMINARY; PRT; 933 AA.  
 AC Q6ZM25;  
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
 DE Hypothetical protein FLJ16561.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI TaxID=9606;  
 RN [1]\_SEQUENCE FROM N.A.  
 RC TISSUE=Synovial membrane tissue;  
 RA Ota T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T.,  
 RA Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,  
 RA Isono Y., Kawai-Hio Y., Saico K., Nishikawa T., Kimura K.,  
 RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,  
 RA Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,  
 RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahara K.,  
 RA Masuho Y., Nagai K., Isogai T.;  
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK131434; BAD18580.1;  
 DR InterPro; IPR000585; Hemopexin.  
 DR InterPro; IPR001212; Somatomedin B.  
 DR Pfam; PF00045; Hemopexin; 2.

DR Pfam; PF01033; Somatomedin B; 1.  
 DR PRINTS; PRO0022; SOMATOMEDINB.  
 DR SMART; SM00120; HX; 2.  
 DR SMART; SM00201; SO; 1.  
 DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN 1.  
 DR PROSITE; PS00524; SOMATOMEDIN B; 1\_--.  
 SQ SEQUENCE 933 AA; 102512 MW; B883773C5BC21A23 CRC64;

Query Match 54.9%; Score 2949.7; DB 2; Length 933;  
 Best local similarity 52.4%; Pred. No. 1e-50;  
 Matches 576; Conservative 0; Mismatches 0; Indels 523; Gaps 2;

QY 1 MANKTPIYLLLLSVEFVIOQVSSQELSCKGRCFESFERGECDCDAQCKKYDKCCPDYE 60  
 DB 1 MANKTPIYLLLLSVEFVIOQVSSQELSCKGRCFESFERGECDCDAQCKKYDKCCPDYE 60  
 QY 61 SFCAE----- 65  
 DB 61 SFCAEVHNPTSPSSKKAPPPGASQTIKSTTKRSKPPNKKTKKVIESERITEHSVS 120  
 QY 66 -----VKDNKNRTKKKPTPKPPVUDE 87  
 DB 121 ENCESSSSSSSSSSSTIRKIKSKNSAANRELQKKLVKDNKNRTKKKPTPKPPVUDE 180  
 QY 88 AGSLDNGDFKVTTPDTSTTOHNKVSSTPKITTAKEINPRPSLPNNSDTSKETSITVNKE 147  
 DB 181 AGSLDNGDFKVTTPDTSTTOHNKVSSTPKITTAKEINPRPSLPNNSDTSKETSITVNKE 240  
 QY 148 TVVETKETTNNKQSTDQGEKITSKETQSIKTSKADLAPTSKVLAKETPKAETTTKG 207  
 DB 241 TVVETKETTNNKQSTDQGEKITSKETQSIKTSKADLAPTSKVLAKETPKAETTTKG 300  
 QY 208 PALTTPKEPTTPPKEPASTTPKEPTTTIKSAPTTKKEPAFTTTKSAPTTKKEPAFTTT 267  
 DB 301 PALTTPKEPTTPPKEPASTTPKEPTTTIKSAPTTKKEPAFTTTKSAPTTKKEPAFTTT 360  
 QY 268 KEPAPTTPKEPAFTTTKKEPAFTTTKSAPTTKKEPAFTTTKKEPAFTTTKKEPAFTTP 327  
 DB 361 KEPAPTTPKE----- 370  
 QY 328 TTPKEPAFTTKKEPAFTTPKKEPAFTTPKKEPAFTTPKKEPAFTTTKESPTTPKE 387  
 DB 371 ----- 370  
 QY 388 PAPTTPKAPTTTKKEPAFTTTKSAPTTPKEPSPTTTKKEPAFTTPKKEPAFTTP 447  
 DB 371 ----- 370  
 QY 448 KEPAPTTPKEPAFTTTKKEPAFTTPKKEPAFTTPKKEPAFTTPKKEPAFTTPKEPA 507  
 DB 371 ----- 370  
 QY 508 PTTPELAPTTPPEPTTPPEPAFTTPKAAAPNTEKEPAFTTPKKEPAFTTPKKEPAFTTP 567  
 DB 371 ----- 370  
 QY 568 KETAPTTPKGTAPTTLKEPAFTTPKKEPAFTTTKKEPTSTSDKPAFTTPKGTAPT 627  
 DB 371 ----- 370  
 QY 628 PKEPAFTTPKKEPAFTTPKGTAPTTLKEPAFTTPKKEPAFTTTKGTSTTSKPAFT 687  
 DB 371 ----- 370  
 QY 688 TPKEAPTTPKKEPAFTTPKKEPAFTTPPEPTTPPEPTTTSEVSTPTTTKKEPTTIHKSPDESPELS 747  
 DB 371 -----PAPTTPPEPTTPPEPTTTSEVSTPTTTKKEPTTIHKSPDESPELS 410  
 QY 748 AEPTPKALENSKPEPGVPTTKTTPAATKPEMTTTAKDKTTERDLRTTPTTTAAAPKMTKET 807  
 DB 411 AEPTPKALENSKPEPGVPTTKTTPAATKPEMTTTAKDKTTERDLRTTPTTTAAAPKMTKET 470  
 QY 808 ATTTEKTTESKITATTQVTTTQDTTPFKLTTLTKTTLAPKVTITTKTITTTIMNKP 867

```

Db      471 ATTTEKTTESKITTATTTQVTSITTTQDTTPFKITTLTKITTLAPKVTTTKITITTEIMNKP 530
Qy      868 EETAKPKDRATNSKATTPPKQPKTKAPKKPTSTKKPKTMVRKPKTTPPRKMTSTMP 927
Db      531 EETAKPKDRATNSKATTPPKQPKTKAPKKPTSTKKPKTMVRKPKTTPPRKMTSTMP 590
Qy      928 LNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSDAGGAGETPHMLLRPHVFMPEVTPD 987
Db      591 LNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSDAGGAGETPHMLLRPHVFMPEVTPD 650
Qy      988 MDYLPVNPQGIINPMLS 1006
Db      651 MDYLPVNPQGIINPMLS 669

RESULT 4
BAD18580
ID BAD18580 PRELIMINARY; PRT; 933 AA.
AC BAD18580;
DT 12-MAY-2004 (TrEMBLrel. 27, Created)
DT 12-MAY-2004 (TrEMBLrel. 27, Last sequence update)
DE CDNA FLJ16561 fis, clone SYN0V4003981, moderately similar to Homo
DE sapiens proteoglycan 4, (megakaryocyte stimulating factor, articular
DE superficial zone protein, camptodactyly, arthropathy, coxa vara,
DE pericarditis syndrome) (PRG4).
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TTSUE=Synovial membrane tissue;
RA Ota T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T.,
RA Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagatsuma A., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K.,
RA Masuko Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK131434; BAD18580.1; -.
SQ SEQUENCE 933 AA; 102512 MW; B8837735BC21A23 CRC64;

Query Match 54.9%; Score 2949.7; DB 2; Length 933;
Best local similarity 52.4%; Pred. No. 1e-50;
Matches 576; Conservative 0; Mismatches 0; Indels 523; Gaps 2;

Qy      1 MAWKTLPIYLLLLSVFVIOQVSSQELSCKGRCFESFERGECDCDAQCKYDKCCPDYE 60
Db      1 MAWKTLPIYLLLLSVFVIOQVSSQELSCKGRCFESFERGECDCDAQCKYDKCCPDYE 60
Qy      61 SFCAE----- 65
Db      61 SFCAEVNTPSPSSKKAPPPSGASQIKSTTKSPKPNKKTKKVIIEEITEHSVS 120
Qy      66 -----VKDNKNRTKKKPTPKPPVDE 87
Db      121 ENQESSSSSSSSSSSTIRKIKSSKNSAANRELQKLKVKDNKNRTKKKPTPKPPVDE 180
Qy      88 AGSLDNGDFKVTTPDSTTQHNKVSIPKITTAKPINRPSLPNSDTSKETSITVKE 147
Db      181 AGSLDNGDFKVTTPDSTTQHNKVSIPKITTAKPINRPSLPNSDTSKETSITVKE 240
Qy      148 TTVEKTKTTTNKQTSIDGKEKTTSAKETSIETSAKDLAPTSKVLAKPTPKAETTTKG 207
Db      241 TTVEKTKTTTNKQTSIDGKEKTTSAKETSIETSAKDLAPTSKVLAKPTPKAETTTKG 300
Qy      208 PALTPKPEPTTPPKPEASTTPKBPPTTTKSAFTTPKPEAPTTKSAFTTPKBPPTTT 267
Db      301 PALTPKPEPTTPPKPEASTTPKBPPTTTKSAFTTPKPEAPTTKSAFTTPKBPPTTT 360

```

```

Qy      268 KEPAFTTPKEPATTITKEPATTITKSAPTTTPKEPATTTPKAPATTTPKEPATTTPKEPTP 327
Db      361 KEPAFTTPKE----- 370
Qy      328 TTTPKEPATTTPKEPATTTPKEPATTAPKAPATTTPKEPATTTPKBPATTTPKBPSTTPKE 387
Db      371 ----- 370
Qy      388 PAPTITKSAPTTITKEPATTITKSAPTTTPKEPSTTTTKEPATTTPKEPATTTPKBPATTTP 447
Db      371 ----- 370
Qy      448 KEPAFTTPKEPATTITKBPATAPKEPATTTPKETAPTTPKCLTTPTPKCLAPTTPKPA 507
Db      371 ----- 370
Qy      508 FTTPBELAFTTPPEPTTTPPEBPATTTPKAAAPNTPKBPATTTPKEPATTTPKEPATTTP 567
Db      371 ----- 370
Qy      568 KETAPTTPKGTAPTITLKEPATTTPKBPAPKELAPTITTKBPTSTTSKBPATTTPKGTAPT 627
Db      371 ----- 370
Qy      628 PKEPATTTPKEPATTTPKGTAPTITLKEPATTTPKBPAPKELAPTITTKBPTSTTSKBPAPT 687
Db      371 ----- 370
Qy      688 TPKETAPTTPKEPATTTPKBPATTTPETPPPTTSVSTPTTKBPTTIHKSPDESTPELS 747
Db      371 -----PAPTTPETPPPTTSVSTPTTKBPTTIHKSPDESTPELS 410
Qy      748 AEPTPKALENSPKERGVPPTTKTAAATKPEMTTAKDKTTERDLRTTPETTTAAAPKMTKET 807
Db      411 AEPTPKALENSPKERGVPPTTKTAAATKPEMTTAKDKTTERDLRTTPETTTAAAPKMTKET 470
Qy      808 ATTTEKTTESKITATTQVTSITTTQDTTPFKITTLTKITTLAPKVTTTKITITTEIMNKP 867
Db      471 ATTTEKTTESKITATTQVTSITTTQDTTPFKITTLTKITTLAPKVTTTKITITTEIMNKP 530
Qy      868 EETAKPKDRATNSKATTPPKQPKTKAPKKPTSTKKPKTMVRKPKTTPPRKMTSTMP 927
Db      531 EETAKPKDRATNSKATTPPKQPKTKAPKKPTSTKKPKTMVRKPKTTPPRKMTSTMP 590
Qy      928 LNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSDAGGAGETPHMLLRPHVFMPEVTPD 987
Db      591 LNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSDAGGAGETPHMLLRPHVFMPEVTPD 650
Qy      988 MDYLPVNPQGIINPMLS 1006
Db      651 MDYLPVNPQGIINPMLS 669

RESULT 5
Q9JMN99
ID Q9JMN99 PRELIMINARY; PRT; 1054 AA.
AC Q9JMN99;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus.
DE Name=PrG4;
DE Mus musculus (Mouse);
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20573856; PubMed=11124536;
RA Ikegawa S., Sano M., Koshizuka Y., Nakamura Y.;
RT "Isolation, characterization and mapping of the mouse and human PRG4
RT (proteoglycan 4) genes.";

```

```

RL Cytogenet. Cell Genet. 90:291-297(2000).
DR EMBL; AB034730; BAA92310.1; -.
DR HSP; P04004; IOC0.
DR MGD; MGI:1891344; Prg4.
DR InterPro; IPR000585; Hemopexin.
DR InterPro; IPR001212; Somatomedin B.
DR Pfam; PF00045; Hemopexin; 2.
DR PRINTS; PF01033; Somatomedin B; 2.
DR PRINTS; PR00022; SOMATOMEDINB.
DR SMART; SM00120; HX; 2.
DR SMART; SM00201; SO; 2.
DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN 1.
DR PROSITE; PS00524; SOMATOMEDIN B; 2.
DR PROSITE; PS00524; SOMATOMEDIN B; 2.
SQ SEQUENCE 1054 AA; 115991 MW; 4FC64BFA42283235 CRC64;

Query Match 42.98; Score 2307.7; DB 2; Length 1054;
Best Local Similarity 40.78; Pred. No. 4.8e-38;
Matches 470; Conservative 37; Mismatches 134; Indels 513; Gaps 20;

QY 1 MAWKTLPIYLLLLSVFVIOQVSSQ----- 25
DB 1 MGWKILPVCLSLLPVVLIOQVSSQDLSSCAGRCGEGYSDATCNDYNCQHYWECPCDF 60
QY 26 -----ELSCKGRCFSEFRGRCDCAQCKYDKCPDYSEFCAE----- 65
DB 61 KRVCSPELSCGRCFSEFARGRECDQSCQYKGCADYDFCEEVHNSTSPSSKTAPT 120
QY 66 ----- 65
DB 121 PAGASTIKSTTKRSPKSPTRTRIKVSEBELTEHSDSENQPSSSSSSSSTIRKIS 180
QY 66 -----VKDNKNRTKKKPTPKPPVVDAGSLDNGDFKVT--TPDSTTQH 109
DB 181 SKNSANRELQKNPNVNDKNNTKKFENPEPPAVDAGSLDNGERFKLTPPPDPPTTEH 240
QY 110 NKVSTSPKLTAKPINRPSLPNSDTSKETSITVNEKETTETTTNKQSTQDQK 169
DB 241 SKVATSPKTTAAKPVTPKPSLAPNSETSKASLASNKETTETTKETTATNKQSSA-SKKK 299
QY 170 TTSAKETOSIETSAKDLAPTSKVLAKPTPKAETITTKGALITPKPEPTTTPKEPASTTP 229
DB 300 TTSVKETRSKAEKTSKDVSEPT----- 321
QY 230 KBPTPTTIKSAPTTPKEPAPTTTKSAPTTPKBPAPTTTKEPAPT-----PKBPAPT 281
DB 322 -----TTPKNSAPTTTKKPVTTTKESKFLPLQPEPT 354
QY 282 TTKEPAPTTTKAPTTKBPAPTTTKKBPAPTTTPKEPAPTTTPKEPAPTTKEPA 341
DB 355 TAKEPPPTTKPEPTTKPEPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 414
QY 342 PTTKPEPAPTAPKAPTTKBPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKSAPT 401
DB 415 PTTKPEPAPTAPKAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 461
QY 402 EPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 461
DB 462 -----EPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 512
QY 462 TTKKAPAPKBPAPTTTPKEPAPTTTPKKTLPPTTPEKAPTTTPKEPAPTTPEELAPT 521
DB 513 ----- 512
QY 522 PPTTTPKEPAPTTPKAAAPNTKBPAPTTTPKEPAPTTTPKEPAPTTTPKGTAPT 581
DB 513 -----TPKEPAPTTTPKEPAPTTTPKEPAPTTTP 538
QY 582 TLKEPAPTTPKKAPKAPKELAPTTTKGPTSTSDKAPTTTPKGTAPTTPKEPAPTT 641
DB 539 --KEPEPTTPKP-----EPTTPKEPAPTTTPKEPAPTTTPKEPAPTT 567
QY 642 TTPKGTAPTTLKEPAPTTTPKPKAPKELAPTTTKGPTSTSDKAPTTTPKGTAPTTPKEP 701
DB 642 TTPKGTAPTTLKEPAPTTTPKPKAPKELAPTTTKGPTSTSDKAPTTTPKGTAPTTPKEP 701

```

Db 568 TTP-----KEPEPTTPKEPEP-----TTRKEPPTTPKEPEPTTPKEPE 606

QY 702 PTPPKAPPTTPPETPTTSEVSTPTTKEFTTIHKSPDESTPELSAEPPTKALENSPKE 761

Db 607 PTPPKKEPTT----- 617

QY 762 PGVPTTKTPAATKPEMTTITAKDKTERDLRTTPTTTAAPKMTKEATATTEKTTESKITA 821

Db 618 -----TSPKT----- 622

QY 822 TTTQVSTTTQDTPPKIITLTKTTILAPKVTTTKTITTTTMMKPBETAKPKDRATNSK 881

Db 623 -----TTLKATTLAPKVTAPE-----EIQNKPEETTPASEDSDDSK 659

QY 882 AT-----TPKP-QKPTKAPKKPTSTKKPKTMPRVRKPTTTPPKMTSTMPELNPTS 932

Db 660 TTLKPQKPTKAPKPTKKPTKAPKKPTSTKKPKT-PKTRKEPTTPAPLKTTSAPELNTTP 718

QY 933 RIAEAMLOTTTRNOTPNSKLVNPKSEDAAGAEHTPHMLRPHVFMPEVTPDMYLP 992

Db 719 --LEVMLPTTIIPKQTPNPEAEVNPDPHEDADGEGEKP-LIPGPPVLPFTAIPGTDLA 775

QY 993 RVENQGIINPMLS 1006

Db 776 GRLNRGININPMS 789

RESULT 6

MUC2\_HUMAN STANDARD; PRT; 5179 AA.

ID MUC2\_HUMAN

AC Q02817; Q14878;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 05-JUL-2004 (Rel. 44, Last annotation update)

DE Mucin 2 precursor (Intestinal mucin 2).

GN Name=MUC2; Synonyms=SMUC;

OS Homo sapiens (Human);

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Intestine;

RX MEDLINE=94132002; PubMed=8300571;

RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;

RT "Molecular cloning of human intestinal mucin (MUC2) cDNA.

RT Identification of the amino terminus and overall sequence similarity

RT to prepro-von Willebrand factor."

RL J. Biol. Chem. 269:2440-2446(1994).

RN [2]

RP SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.

RC TISSUE=Colon;

RX MEDLINE=93016075; PubMed=1400449;

RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe B.-M., Lagace R.E.,

RA Kim Y.S.;

RT "The human MUC2 intestinal mucin has cysteine-rich subdomains located

RT both upstream and downstream of its central repetitive region."

RL J. Biol. Chem. 267:21375-21383(1992).

RN [3]

RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.

RX MEDLINE=91358717; PubMed=1885763;

RA Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,

RA Petersen G.M., Kim Y.S.;

RT "MUC-2 human small intestinal mucin gene structure. Repeated arrays

RT and polymorphism."

RL J. Clin. Invest. 88:1005-1013(1991).

CC -!- FUNCTION: Coats the epithelia of the intestines, airways, and

CC other mucus membrane-containing organs. Thought to provide a

CC protective, lubricating barrier against particles and infectious

CC agents at mucosal surfaces.

CC -!- SUBUNIT: Multimeric.

CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- TISSUE SPECIFICITY: Colon, small intestine, colonic tumors,







RA	Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,	
RA	George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,	
RA	Paclet J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,	
RA	Svirskas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,	
RA	Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.,	
RT	"Finishing a whole-genome shotgun: release 3 of the <i>Drosophila</i>	
RT	melanogaster euchromatic genome sequence."	
RL	Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).	
RN	[3]	
RN	SEQUENCE FROM N.A.	
RX	MEDLINE=22426070; PubMed=12537573;	
RA	Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,	
RA	Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,	
RA	Ashburner M., Calniker S.E.;	
RT	"The transposable elements of the <i>Drosophila melanogaster</i> euchromatin:	
RT	a genomic perspective."	
RL	Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).	
RN	[4]	
RN	SEQUENCE FROM N.A.	
RX	MEDLINE=22426069; PubMed=12537572;	
RA	Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,	
RA	Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,	
RA	Smith C.D., Tupy J.L., Whitfield E.J., Bayraktarov L., Berman B.P.,	
RA	Battencourt B.R., Celisner S.E., de Grey A.D., Drysdale R.A.,	
RA	Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,	
RA	Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,	
RA	Lewis S.E.;	
RT	"Annotation of the <i>Drosophila melanogaster</i> euchromatic genome: a	
RT	systematic review."	
RL	Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).	
RN	[5]	
RN	SEQUENCE FROM N.A.	
RG	FLYBASE;	
RL	Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.	
RN	[6]	
RN	SEQUENCE FROM N.A.	
RG	FLYBASE;	
RL	Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AE003575; AAF50957.3; -.	
DR	FLYBase; FBgn0003372; Sgs1.	
SQ	SEQUENCE 1225 AA; 127735 MW; A99AF9D404C79FA CRC64;	
	Query Match 24.3%; Score 1306.5; DB 2; Length 1225;	
	Best Local Similarity 27.2%; Pred. No. 2.6e-18;	
	Matches 344; Conservative 117; Mismatches 42; Indels 385; Gaps 45;	
Qy	10 LLLLSVFIQOVS-----SQBL--SKGRCFESFERGECDCACKKYD 53	
Db	6 IFLVSIILIQVKNVKNYDWSQDGSSEHIIFCGGDTIYDPVPQCTDS----- 59	
Qy	54 KCCPDYSEFCAEVKDNKNRKKXP---TPKPPVVDEAGSGLDNGDFKVTTPDTSTTQHN 110	
Db	60 -----NPTTKPRQTKRP-----KSTRITTKRTKRP 86	
Qy	111 KVST-----SPKITAKPINRPSLPNSDTSKETSITVKNKTVE-----T 152	
Db	87 RKTKTKTKRATK 146	
Qy	153 KETTTTKQTSD-----GKEKTTSAKETQSIKTSKADIAPTSKVLAKPTPKAE 202	
Db	147 DRTIVNTDWTPLCTDTPCTCSESSESTAIPSSPCIDTSTVITPSQCTQTTPPTCS 206	
Qy	203 -----TTTKGPALTTPKEPTTTPKEPASTTPKEPTTPTIKS 239	
Db	207 TQGTQTPCTCAQTTTPRSTTTTTSRPTTTTTPRSTTTTTSRPTTTTTPRSTTTTTPR 266	
Qy	240 APTTPKEPAPTTKSAKAP---TTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTKSAPTT 297	
Db	267 PTTTTPRCTTITSICAPITTPRSATTTTSRPTTTTTPRSTTTTTCSPTRTPRSTTTT 326	
Qy	298 -PKEPAPTP-----KKPAPTPKE-----PAPTPKE-----PTPTT 329	
Db	327 STSRPTTTTPRCTTTPSTTCTCTPRPTTTTTPRSTTTTCTSGPTTTPRSTTTCTSGPTT 386	



OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OX Ephydroidea; Drosophilidae; Drosophila.  
RN NCBI\_TaxID=7227;  
RX [1]  
RA MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Anantides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.P.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.H., Blazer R.G., Champagne M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,  
RA April J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Folsler C., Gabrieli A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwan C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong M., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RA "The genome sequence of *Drosophila melanogaster*."  
RA Science 287:2185-2195(2000).  
RN [2]  
RX SEQUENCE FROM N.A.  
RX MEDLINE=22426065; PubMed=12537568;  
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,  
RA Patel S., Adams M., Champagne M., Dugan S.P., Frise E., Hodgson A.,  
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,  
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
RA Weinstock G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.,  
RT "Finishing a whole-genome shotgun: release 3 of the *Drosophila*  
RT *melanogaster* euchromatic genome sequence."  
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
RN [3]  
RX SEQUENCE FROM N.A.  
RX MEDLINE=22426070; PubMed=12537573;  
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,  
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
RA Ashburner M., Celniker S.E.;  
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
RT a genomics perspective."  
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
RN [4]  
RX SEQUENCE FROM N.A.  
RX MEDLINE=22426069; PubMed=12537572;  
RA Mitra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradscky P., Huang Y., Kaminker J.S., Millburn G.H., Enoch S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,

RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
RT systematic review."  
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).  
RN [5]  
RX SEQUENCE FROM N.A.  
RX FLYBASE;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RX SEQUENCE FROM N.A.  
RX FLYBASE;  
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
RL EMBL: AE003626; AAS64673.1;  
DR InterPro; IPR011009; Kinase\_like.  
FT NON\_TER 1  
SQ SEQUENCE 1761 AA; 197521 MW; 62A5E16E6241B3F4 CRC64;  
Query Match 24.2%; Score 1299.5; DB 2; Length 1761;  
Best Local Similarity 24.2%; Pred. No. 5.5e-18;  
Matches 441; Conservative 114; Mismatches 361; Indels 905; Gaps 82;  
QY 12 LLLSVFV-IQQVSSQBLSCCKGRCFESFERGECDCDAQCKKYDKC-----CPDYVESF 62  
Db 1 VLVSFLWALYHSANAALG-----SEIKGDESTTESAEKYDYCWINPLPGCP----- 52  
QY 63 CAEVKONKNRTKKKTPKPPVVDGAGSLDNGDFKVTTPDTSSTOHNVKYSTSPKITT-- 120  
Db 53 -----FGPGDGNSTTKISTLATTKSTVTITSEETLTK 85  
QY 121 -----AKPINRPSLPNSDTSKETSITVAKETIVETKNTKQSTGKE-----K 169  
Db 86 ITTIKSTAKPTQKTNEPTT-----KITTPKATTI--KSTATTARATEAPKTEOTTLR 138  
QY 170 TTSAKETQ-----SIEKTSADLAPTSKVLAKPPTKPAETTKGPAITTKPTKEPT- 217  
Db 139 TTIKSTSELITLTKTTIKSAETKSHNPTTKASTT-----LRITEPTTKKSTAK 193  
QY 218 -----PTTPKEPASTTPKEPT-----PTTIKSA- 245  
Db 194 TTREPTTKRETTERTTQEPSTSKTTTHTTAAEPATKKTHTTPTQKSTTLRITEPTTK 253  
QY 246 EPAPTTPKSAPTTPKEPAPTTPKEPA----- 271  
Db 254 SSTAKITRE-PTTKGETTERTTKEPSTSKTTTHTTAAEPATKKTHTTPTQKSTTLRITE 312  
QY 272 -----PTTPKEPAPTTPKEPA----- 287  
Db 313 EPTTRKSSSTARRPTTKRETTERTTQEPSTSKTTTHTTAAEPATKKTHTTPTQKSTT 372  
QY 289 -----PTTIKSA-----PTTPKEPAPTTPKKA- 313  
Db 373 LRITEPTTRKSSSTARRPTTKRETTERTTKEPSTSKTTTHTTAAEPATKKTHTTPT 432  
QY 314 PK-----EPA-----PTTPKEPTTPKEPA-----PTTPKEPA- 341  
Db 433 QKSTTLRITEPTTRKSSSTARRPTTKRETTERTTQEPSTSKTTTHTTAAEPATKKT 492  
QY 342 -----PTTPK-----EP-----APTA 352  
Db 493 HEPTTQKSTTLRITEPTTRKSSSTARRPTTKRETTERTTQEPSTSKTTTHTTAAEP 552  
QY 353 PKKA-PTTPK-----EPA-----PTTPKEPAPTTPKEPS- 381  
Db 553 TKKTHTTPTQKSTTLRITEPTTRKSSSTARRPTTKRETTERTTQEPSTSKTTTHT 612  
QY 382 -----PTTPKEPA-----PTTIKSA-----PT- 398  
Db 613 TAEPAKKTHTTPTQKSTTLRITEPTTRKSSSTARRPTTKRETTERTTQEPSTSKT 672  
QY 399 -----TTKEPA-----PTTIKSA-----PTTPKPSPTTKPEP 426  
Db 673 THTTAEPAKKTHTTPTQKSTTLRITEPTTRKSSSTARRPTTKRETTERTTPTKEP 732

QY 427 A-----PTTPK-----BPA-----PTTPKKAPT 445  
 Db 733 TTRKTTTHKTEPTTKTTKTEPTTKSTLTKTEPTTRKSTTKTTEPTTKRKTTR 792  
 QY 446 TPEKA-----PTTPKEPAPTTPKPA-----PTAPK-----EPA----- 475  
 Db 793 TTKEPTTRKTTTHKTEPTTK-----TTTKTTHEPTTKSTLTKTEPTTRKSTTK 848  
 QY 476 -----PTTPKETA-----PTTPKKLT-----PTT-P 495  
 Db 849 TTREPTTKETTERTTQEPSTSKTTTHETAEPATKTTHEPTTQKSTTLRITTEPTTRK 908  
 QY 496 EKLAPTTPKEPAPTTPPEAPT-----PTTPE-----EPA----- 531  
 Db 909 SSTAKTTRE-----PTTKRETTERTTKETTRKTTTHKTEPTTKTTHEPTTKSTLTKP 965  
 QY 532 ---PTTPKAAAPTTPKEPAPTTPKEAPTTPKEPA-----PTTPKETA----- 571  
 Db 966 TTEPTTRKSTTKTRE-----PTTKRKTERTKEPTTRKTTTHKTEPTTKTTKTTTH 1023  
 QY 572 -PTTPKGTA-----PTTLKEPAPTTPKKPAKELA-----PTTKGPTSTTSKPA- 616  
 Db 1024 EPTTKSTLTKPTEPTTRKSTTKTTEPTTRKRVITERTREPTTKSTTKTTHETAEPAT 1083  
 QY 617 -----PTTPKGTA-----PTTPKEPAPTTPKEPA-----PTTPKGTAAPTTL 652  
 Db 1084 KKTHEPTTQKSTLRAITEPTTRKSTAKTTREPTTRKRETTERTKEPTTRKTTTHKTT 1143  
 QY 653 KEPA-----PTTPK-----KPAPKELAPTTTKGPTSTSDKAPTTPKETAAPTTPKEPA 701  
 Db 1144 EPTTKTTHEPTTKSTLTKPTEE-----PTTRK-----TSTTKTTEPTTRKTTERTKEPT 1198  
 QY 702 -----PTTPK-----KP-----APTTPETPPPTTSEV 723  
 Db 1199 TRKTTTHKTEPTTKTTTKTTHEPTTKSTLTKPTEPTTRKSTTKTTEPTTRKRV 1258  
 QY 724 STPTTKPE-----TTTHKSDE-----STPELSABEPTPKALENSKEPGVPTTKPAAT 773  
 Db 1259 TTERTEPTTRKTTTHKTEPTTKTTTKTTHEPTTKK-STLTKPTEPTTRKTTST 1317  
 QY 774 KPMTTITAKDTTERDLR-----KVTITTTIMNKP-----TTPETTTAAP-- 801  
 Db 1318 KTREPTTKKAVTERTREPTTRKTTTHKTEPTTKTTTKTTKTTHEPTTKSTLTKPTE 1377  
 QY 802 -----KMTKEPAT-----TTEKTESKITATTTQVSTTTQDTPPTPKITLTKT----- 844  
 Db 1378 EPTTRKSTTKTTEPTTRKRVITERTREPTTRKTT--THKTEPTTKTTTKTTKTTHEP 1435  
 QY 845 -----TTLAP-----KVTITTK-----KTITTTIMNKP-----EETAKPKDRA 877  
 Db 1436 TTKKSTLTKPTEPTTRKSTTKTTEPTTRKRVITERTREPTTRKTTTHKTTHEPTTK 1495  
 QY 878 TNSKATTPKQ-----KP-----TKAPKKPT----- 898  
 Db 1496 TTKTTHEPTTKSTLTKTEPTTRKSTTKATREPTTRKRVITERTREPTTRKTTAH 1555  
 QY 899 -----STKKPKTMRVVRKPTTTPTRKMTSTWPELNPSTRIEAM 930  
 Db 1556 KTTEPTTKTTTKTTKTTHEPTTKSTTL-----KTEPTTRKSTTKTTEPTTRKSTTS 1611  
 QY 939 LQTT-----TRPQNTNSKLNVNPKSEDAGGAGET----- 970  
 Db 1612 VKTTADTTKRTTAEMSTTQEPSTSVETTTNSQNSQNTTSTTTERQHVHHHHHHIYH 1671  
 QY 971 -----PHMLLRPH----- 978  
 Db 1672 KPADLGSILPLDPLPLPLPLPLBPLPLPLPLPLPLPLPLPLPLPLPLPLPL 1731  
 QY 979 ---VEMPEVT-PMDYLPVFN 996  
 Db 1732 TAILSLPEISLPLPLPLPLPN 1752

RESULT 9  
 AAS64673 PRELIMINARY; PRT; 1761 AA.  
 AC AAS64673;  
 DT 01-APR-2004 (TREMBlrel. 27, Created)  
 DT 01-APR-2004 (TREMBlrel. 27, Last sequence update)  
 DT 01-APR-2004 (TREMBlrel. 27, Last annotation update)  
 DE CG33300-PA (Fragment).  
 DE CG33300.  
 GN Drosophila melanogaster (Fruit fly).  
 OS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang O., Chen L.X.,  
 RA Brandon R.C., Rogers J.H., Blazek R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,  
 RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Heston D.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hottel N.L., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Rainert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426065; PubMed=12537568;  
 RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,  
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
 RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,  
 RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
 RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
 RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
 RT "Finishing a whole-genome shotgun: release 3 of the Drosophila  
 RT melanogaster euchromatic genome sequence.";  
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426069; PubMed=12537572;  
 RA Miera S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradscky P., Huang Y., Celniker J.S., Millburn G.H., Prochnik S.E.,





```
QY 715 -----TPPPTT-----SEVSTPTTKBPTTIHKSP-----739
D 2393 DTTMSSASTPESTTGGTTRTTPRTPTDSTMSASTPESTTGGTTRTTPRTSTES 2452
QY 740 -----DESIPELSAEP-----TPKALENSPKPEP-762
D 2453 TDTMSSASTPESTTGGTTRTTPRTPTSTESTDTMSSASTPESTTGGTTRTTPRT 2512
QY 763 ---GVPTTKTAAKPEWTTAKKTIERDLRTPTETTTAAKPKMTKETATT-----810
D 2513 STSTDTMTAST-PEPSIT-PDTTTRTTPRTPTDTMSSASTPESTTGGTTRTTP 2570
QY 811 -----TBKTTESKITATTQVTSSTTQDTPPKITLLKTTTLA 848
D 2571 RPTPTDTMSSASTPESTTGGTTRTTPRTPTSTESTDTMSSASTPESTTGGTTRT 2630
QY 849 P-----KVTTTKTITTEIMNKPEETAKPKDRATN 879
D 2631 PTPPTPTDSTMSSMSASATPGTTRTTPRTPTSTESTDTMSSASTPESTTGGTTRT 2690
QY 880 -----SKATTPKPO-----KPTK-----892
D 2691 PTPPTSTDTMSSMSASTPESTTGGTTRTTPRTPTDSTMSSMSASTTGGTTRT 2750
QY 893 APKPTPTK-----KPTMP---RVRKPKTTPTRKMTSTM-----PELNP 930
D 2751 TPTPTSTESTDTMSSASTPESTTGGTTRTTPRTPTDSTMSSMSASTPESTT 2810
QY 931 -TSRIAEMLQTTTRNPQTNPKLVNPKSEDAGAGBGTPHMLRPHVFMPEVTP 986
D 2811 GTTR-----TTPTPTPTDSTMSSMSAS-----TPE-----PSTTP 2843

RESULT 11
Q6SSE6 PRELIMINARY; PRT; 3409 AA.
AC Q6SSE6;
DT 05-JUL-2004 (TEMBLrel. 27, Created)
DT 05-JUL-2004 (TEMBLrel. 27, Last sequence update)
DE plus agglutinin.
GN Name-SAG1;
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RA Ferris P.J., Waffenschmidt S., Umen J.G., Ishida K., Kubo T., Lau J.,
RA Goodenough U.W.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY450930; AAS07044.1;
DR InterPro; IPR002951; Atrophin.
DR InterPro; IPR003882; Pibcil_extensin.
DR PRINTS; PRO1222; ATROPHIN.
DR PRINTS; PRO1218; PSTLEXTENSIN.
SQ SEQUENCE 3409 AA; 336045 MW; 7FE87633EDD6631F CRC64;

Query Match 21.3%; Score 1141.9; DB 2; Length 3409;
Best Local Similarity 17.9%; Pred. No. 1.5e-14;
Matches 303; Conservative 127; Mismatches 365; Indels 898; Gaps 47;

QY 76 KKPTPPVVDAGSLDNGDFKVTTPDT---STQHKNKVSPTTKTAKPINPRESLPP 132
D 574 RAPRESPP-----FHPSPDPPASSVPPSPKPPSPAPSPAPSP 620
QY 133 NSDTSKETSIVNKETTVEKTTTTNKQTSIDGKEKTSKETSIAKDLAPTSK 192
D 621 -----PESP 624
QY 193 VLAKPTPKAEITTKGPAULTTPKEPTPTTKPEAST-----TPKEPTPTTIKSAFTTPK 245
D 193 -----TPKEPTPTTIKSAFTTPK 245
```

```
Db 625 APPSPAPPS-----PAPSPQPPSPVPPOPPSPVPSPKPPSPAPSPVPSPAPSPSPA 678
QY 246 EPAPTITTKSAFTTTKEPAPTITTKPAPTTPK-----EPAPTITTKPAPTITTKSAFTTPKEP 301
D 679 PPSAPENPAPPSAPPLPPSPPEPSPAPSPPEPSPAPSPPEPSPAPSPPEPSPAPSPAPSPAPP 738
QY 302 APITTPKKPAPT-----TPKEPAPTTPKSPPTTTPKEPAPTTPKEPAPTTPKE 347
D 739 APPSPPEPSPAPSPPEPSPAPSPPEPSPAPSPPEPSPAPSPPEPSPAPSPPEPSPAPSPSEP 798
QY 348 PAPT-----APKKPAPTTPKEPAPTTPKEPAPT-----375
D 799 PPSAPSPAPSPPEPSPAPSPPEPSPAPSPPEPSPAPSPPEPSPAPSPPEPSPAPSPVPPS 858
QY 376 -TTKEPSPPTTKPAPTITTKSAFTTTKEPAPTITTKSAFTTTKEPSPPTTKPAPTITTKPAPT 434
D 859 PAPPSPAPSPPEPSPAPSPPEPSPAPSPPEPSPAPSPPEPSPAPSPPEPSPAPSPAPP 917
QY 435 AP-----TTPKKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKKPAPTAPK 472
D 918 SPPEPSPPEPSPAPSPPEPSPAPSPPEPSPAPSPPEPSPAPSPPEPSPAPSPPEPSPAPSPA 977
QY 473 -----EPAPTTPKETAPTTPKKLTPTTEKLAAPTPE---KEAPTTPPEE 513
D 978 PPSDPSPAPSPDPSPAPSPAPSPPEPSPPEPSPPEPSPPEPSPPEPSPPEPSPAPSPPPP 1036
QY 514 LAFT-----TPEEPTTTPKEPAPTTPKAAAPN-----TPKEPAPTTPK 552
D 1037 TPTSPSPSPPEPSPPEPSPAPSPPEPSPAPSPPEPSPAPSPPEPSPAPSPPEPSPAPSPA 1096
QY 553 EPAPTTPK-----EPAPTTPKETAPTTPKGTAPTTL-----583
D 1097 PPSAPSPPEPSPAPSPPEPSPAPSPPEPSPAPSPPEPSPAPSPPEPSPAPSPPEPSPAPSPA 1156
QY 584 -----KEPAPTTPKKPAPKELAPT-----TKEPT 608
D 1157 SPPEVPTTPSPAPSPAPSPPEPSPPEPSPPEPSPAPSPAPSPAPSPAPSPPEPSPPEP 1216
QY 609 STTSKAPAPTTPKGTAPTTPKEPAPTTPKGTAPTTPKGTAPTTPKGTAPTTPKGTAPTTPK 668
D 1217 SPEPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAP 1271
QY 669 APT-ITKGTSTTSKAPAPTTPKEPAPTTPKEPAPTTPKGTAPTTPKGTAPTTPKGTAPTTP 726
D 1272 SPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAP 1328
QY 727 TTTKEPTTIHKSPDESTPELSAETPKALEN-----SPKEPGV-764
D 1329 AAALPPLPSPAPPLPVPFASPAPSPPLRPQOTPAMPSPAPSPAPSPAPSPAPSPAPSPAP 1388
QY 765 ---PTTKTP-----AATKE-----776
D 1389 PPPPTPTPLAPLPDCTLLAQALLSIDAANSVVFVSAGLFTSVARSTPELLASFT 1448
QY 777 -----MTTTL-----781
D 1449 VCSCLTATSLVGGSSRGNNNTGSSSGNYISNGGDAAIQRIAGNSTDRGSSSGSG 1508
QY 782 -----KDKTTERDL-----RTTP-----794
D 1509 GSSSGPGCTETAAEWAVDQDGYQLQSLGGVYTRTVVDRTPPVSNGNVTLSANRI 1568
QY 795 -----ETTTAAPKMT--804
D 1569 KQEPSAVGEASINALGSKQAMLLTISFSEVPAPFAPPAASLIVTGALVAEWAAADKMTFY 1628
QY 805 -----KETATTTEKTTESKIT-----820
D 1629 VLAMTLPAELVATAAGSSSSGTSRSGNGNGTAAAAAAPPAGTTGRRRALQQQAAA 1688
QY 821 -----820
D 1689 PPPFASGSSSSLSGAANTANQQORHVHLLPATAYADAARNPGRNDLSLSVELTDNAVASP 1748
```

```

QY 821 -----ATTQVSTTTQDTPFKITLTKTTTLAPKVT----- 852
Db 1749 AVGEALATTARVTAATYP-----AVAATTTLVAASSSFAQAIKAGSLIQSYHIQM 1801
QY 853 -----TTKKTITTE----- 862
Db 1802 LTMSLYLASRGVGREYGEYAVFKYAVLGVKGNLGPAAEAMPTNEKEVTAASQARQVGD 1861
QY 863 -----IMNKPEETAKPKRATNSKATTPKPKP----- 890
Db 1862 LMPIGNDLLGSSNITASSSSSSSSNSP-PRRPPPPAAGSTGLLFSNADASPPPLA 1920
QY 891 -----TKAPKKTSTKKPKTMVRKPKTTPPRKMTST--MPELN- 929
Db 1921 VATPALPAFLPSTIAAAATAAP-----PRLPSP--PPPAVGSGTGVLPRLHL 1965
QY 930 -----PSTRAEAMLQTTTRNQTPNSKLYE----- 955
Db 1966 MOQMLQPPAAVAAPPPPPASSALVLPSPPPPPPPSOLLIOQASATVSDMDLLYL 2025
QY 956 -----VNPKGEDAGGARGETPHMLLRPHVFMPEVTPDMDYL 991
Db 2026 VVAAMLITAVAGRLIAAIVLYLLVSP-----EHPFLA-----F 2060
QY 992 PR----- 993
Db 2061 PRELTTIAGLILVALTFYSQALGSPAADWHGSRTAAYCVLITIAVVFYAAFLWMLALARA 2120
QY 994 --VPOGILINPM 1004
Db 2121 WMVP-QFTLVEPM 2132

RESULT 12
AAS07044 PRELIMINARY; PRT; 3409 AA.
AC AAS07044;
DT 02-MAR-2004 (TEMBLrel. 27, Created)
DT 02-MAR-2004 (TEMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TEMBLrel. 27, Last annotation update)
DE Plus agglutinin.
GN SAGI.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OC NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RA Ferris P.J., Wafenschmidt S., Umen J.G., Ishida K., Kubo T., Lau J.,
RA Goodenough U.W.;
RT "Plus and Minus Sexual Agglutinins from Chlamydomonas reinhardtii.";
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY450930; AAS07044.1;
SQ SEQUENCE 3409 AA; 336045 MW; 7FE87633EDD6631F CRC64;

Query Match 21.3%; Score 1141.9; DB 2; Length 3409;
Best Local Similarity 17.9%; Pred. No. 1.5e-14;
Matches 303; Conservative 127; Mismatches 365; Indels 898; Gaps 47;

QY 76 KKPTPKPVVDEAGSLDNGDFKVTTPDT---STQHKNKYSTPKITTAKEINPRPSLPP 132
Db 574 RAPRPSPP-----FHPSPDSPASSVPPSPPEPKPSAPPSPAPPSPPP 620
QY 133 NSDTSKETSILVFNKETTIVETKETTNNKQTSIDGKEKTSAKETQSIKTSAKDLAPTSK 192
Db 621 -----PPSP 624
QY 193 VLAKPTKPAETTKGPAALTTPKEPTTPKPEAST-----TPKEPTTIKSAPTPK 245
Db 625 APPSPAPPS-----PAPSPQPPSPVPPSPVPPSPKPPSPAPSPVPPSPAPSPA 678
QY 246 EPAPTTIKSAPTTPKEAPTTTKAPATPK-----EPAPTTTKAPATTTKSAPTPKPE 301

```

```

Db 679 PPSAPPNPAPSPAPPLPPSPSPSPAPSPSPSPAPSPSPAPSPAPSPAPSPAPSPAPP 738
QY 302 APTTPKKAPAT-----TPKEPATTTKEPTTTTPKEPATTTKEPATTTKEPATTTKE 347
Db 739 APPSPSPAPSPSPSPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 798
QY 348 RAPT-----APKKPATTPKEPATTTKEPATTT----- 375
Db 799 PPSAPPSPAPSPSPSPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 858
QY 376 -TTKBPSPPTTPKEPATTTKSAPTTTKEPATTTKSAPTTTKSAPTTTKSAPTTTKSAPT 434
Db 859 PAPPSPAPSPSPSPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 917
QY 435 AP-----TPKKEPATTTKEPATTTKEPATTTKEPATTTKEPATTTKEPATTTK 472
Db 918 SPEPSPSPSPSPSPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 977
QY 473 -----EPATTPKEPATTTKPKLTPTTPEKLAPTTPE--KPAPTTPEE 513
Db 978 PPSPPSPSPAPSPSPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 1036
QY 514 LAPT-----TPPEPTTTDEEPATTPPKAAAN-----TPKEPATTTPK 552
Db 1037 TPPTSPPSPSPSPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 1096
QY 553 EPAPITPK-----EPATTPKEPATTTKGTAPTTL----- 583
Db 1097 PPSAPPSPSPSPAPSPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 1156
QY 584 -----KEPATTTKPKAPKELAPTT-----TKEPT 608
Db 1157 SPBPVPTPPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 1216
QY 609 STTSKAPATTPKGTAPTTKEPATTTKEPATTTKEPATTTKEPATTTKEPATTTKEPAT 668
Db 1217 SPEPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 1271
QY 669 APT-TTKGPTSTSDKAPATTPKETAATTPKEPATTTKEPATTTKEPATTTKEPATTT 726
Db 1272 SPAPPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 1328
QY 727 TTTKEPTTIHKSPDESTEPLSAEPTKALEN-----SPKEPGV- 764
Db 1329 AALPLPSPAPPLPVPPASAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 1388
QY 765 --PTTKTP-----AATKPE----- 776
Db 1389 PPPTPTPLAPLPDCTLLAQALLSIPDAANSFVVSAGLFTSVAPSTPPELLASFCT 1448
QY 777 -----MTTFA----- 781
Db 1449 VCSQLTATATAISLVGSSRRNNNGSSGGNYSGNGGDAALQRIAPAGNSTDRGSSSGS 1508
QY 782 -----KDKTTERDL-----RTTP----- 794
Db 1509 GSSSWGPGTETAAEAWADVQDGYQLSLGIGVYTRTVVVDRTPPSVGNVLSANRI 1568
QY 795 -----ETTTAAPKMT-- 804
Db 1569 KQEPSAVGEASLNALGSKQAMLLTISFSEVPAPFADPAASLIVTGALVAEWAAADKMTFY 1628
QY 805 -----KETATTTTEKTESKIT----- 820
Db 1629 VLAMTLPAELVATAAGSSSSSGTSRSGNGNGTAAAAAAPPAGTTGRRRALQQAAA 1698
QY 821 ----- 820
Db 1689 PPPPASGSSSSLSGAATANQQORHVHLLPATVADAARNDLSILVELTONAVASP 1748
QY 821 -----ATTQVSTTTQDTPFKITLTKTTTLAPKVT----- 852

```

Db 1749 AVGEALATTARVTAATYP-----AVAATTTLVAAASSSFAQAIRAKGSLLOGSYHIQM 1801  
 QY 853 -----TUKTKLITTE----- 862  
 Db 1802 LTMSLYLASRGVGREYGEYAVEFKYAVLVKGNLGPAAEAMPTEKEVTAABQARQVGGD 1861  
 QY 863 -----INMKPETHAKPDRAVNSKATPKQKP----- 890  
 Db 1862 LWPIGNDLLGSSNTTASGSSGSSSSNSP-PRPPPPPPAAGSTGLLFSNADASPPLA 1920  
 QY 891 -----TKAPKPTSTKKEKTPRVKPKTTTPKRMST--MPENL- 929  
 Db 1921 VATPALPAPLPSTTAAATAAP-----PRLPSP---PPPAVGSSTGVLPRRHL 1965  
 QY 930 -----PTSRIBAMLQTTTRPNQTPNSKIVE----- 955  
 Db 1966 MQQWLQPPAAVAAPPPPASSSALVLQSPPPPPPSQLLIQASATYVSDMQDLLYTL 2025  
 QY 956 -----VNPKSDAGGAEGETHMLLRPHVFMPEVTPDMDYL 991  
 Db 2026 VVAAMLITVAAGRLIAAVALYRLIVSPE-----PPFLA-----F 2060  
 QY 992 PR----- 993  
 Db 2061 PRETTIAGILVALTFYSMALGGPAAADWHGSRTAAYCVLTIADVPPYAAFLWMLALARA 2120  
 QY 994 --VPNQGIINPM 1004  
 Db 2121 WMVP-QFTLVEPM 2132

RESULT 13  
 SLPI\_CLOTH  
 ID SLPI\_CLOTH STANDARD; PRT; 1664 AA.  
 AC Q06852;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Cell surface glycoprotein 1 precursor (Outer layer protein B) (S-layer protein 1).  
 DE protein 1).  
 GN Name=olpB;  
 OS Clostridium thermocellum.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1515;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCIB 10682;  
 RX MEDLINE=93209311; PubMed=8458832;  
 RA Fujino T., Beguin P., Aubert J.-P.;  
 RT "Organization of a Clostridium thermocellum gene cluster encoding the  
 RT cellulosomal scaffolding protein cipa and a protein possibly involved  
 RT in attachment of the cellulosome to the cell surface.";  
 RL J. Bacteriol. 175:1891-1899(1993).  
 CC - SUBUNIT: Assembled into mono-layered crystalline arrays.  
 CC - SUBCELLULAR LOCATION: Cell wall.  
 CC - SIMILARITY: Contains 4 S-layer homology (SLH) domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X67506; CA447841.1; -  
 DR PIR; T18262; T18262.  
 DR InterPro; IPR008965; Cellul\_bind.  
 DR InterPro; IPR001119; SLH.  
 DR Pfam; PF003195; SLH; 3.  
 DR PROSITE; PS01072; SLH DOMAIN; 2.  
 KW Cell wall; Repeat; S-layer; Signal.

SIGNAL 1 28 Potential.  
 FT CHAIN 29 1664 Cell surface glycoprotein 1.  
 FT DOMAIN 36 763 4 X 156 AA approximate repeats.  
 FT REPEAT 36 191 1.  
 FT REPEAT 207 363 2.  
 FT REPEAT 409 565 3.  
 FT REPEAT 607 763 4.  
 FT DOMAIN 771 1377 approximate tandem repeats of T-P-S-D-E-P.  
 FT DOMAIN 1378 1449 Gly/Pro/Ser/Thr-rich.  
 FT DOMAIN 1453 1494 SLH 1 (incomplete).  
 FT DOMAIN 1495 1565 SLH 2.  
 FT DOMAIN 1566 1625 SLH 3.  
 FT DOMAIN 1626 1646 SLH 4 (incomplete).  
 SQ SEQUENCE 1664 AA; 178194 MW; 5F396695BA9PE74B CRC64;

Query Match 20.6%; Score 1108.6; DB 1; Length 1664;  
 Best Local Similarity 21.0%; Pred. No. 2.9e-14;  
 Matches 359; Conservative 118; Mismatches 371; Indels 864; Gaps 69;

QY 4 KTLPIYLLLLLSVFV-----IQOVSSQE 26  
 Db 6 KVLISILLTLLIISTTSVNMSPAEATPSIEMVLDKTEVHVGDVITATIKVNNIRKLAGYQ 65  
 QY 27 LSKC-----GRCFSEFERGECDDAQCKYDKCCPDYSEFCAEVKDKNKR 73  
 Db 66 LNIKFDPEVLQVPDPATGEEFT-----DKSMP-----VNR 95  
 QY 74 ----TKKXTP-----KPPVDE----- 87  
 Db 96 VLLTNSKYGTPVAGNDIKGIINFATGVNNLTAYKSGIDEHTGIGIFKVLKQNT 155  
 QY 88 -----AGSGLDNGD-----FKVTTPTDTSTQHNKSTSPKITTAKPINR 127  
 Db 156 SIRFEDTSLMPGALSIGTSLFDWDAETITGVEVIQPDLI-----VVEAEPL----- 200  
 QY 128 PSIPNSDTSKETSUTVNKETT-----VE 151  
 Db 201 -----KQASVALELDKTKVKVGDIIITATIKIENMKNFAGYQLNIKYDPTMLEAIE 250  
 QY 152 TKETTITNKOT-----STDGKEKTS-----AKB 175  
 Db 251 LETGSAIAKRTWPVTGGTVLQSDNFKTAVANDVGAGIINFABAYSNTLYKRETGAEE 310  
 QY 176 TQSIKTSAKDLAPTSKVLAKPTPKAETTT-----KGPALTPP----- 213  
 Db 311 TGIIGKIGFVLKAGSTAI-----RFEDTTMPGALGTYMFDWYGENIKGYSVQPGEI 365  
 QY 214 -----KEPTPT-TPKEPASTTPKEPTPT----- 235  
 Db 366 VAEGEPGEETPEEPVPTETPVDPTFTVTEEPVPSLPSYVIMELDKTKVKVGDIIIT 425  
 QY 236 -----TIKSAPT----- 242  
 Db 426 IKIENMKNFAGYQLNIKYDPTMLEAIELETGSAIAKRTWPVTGGTVLQSDNFKTAVAN 485  
 QY 243 ----- 242  
 Db 486 DVGAGIINFABAYSNTLYKRETGVAAETGIIGKIGFVLKAGSTAIRFEDTTMPGALG 545  
 QY 243 -----TPKEPASTTPKAPTTPKEPAPTTPKEAPT----- 273  
 Db 546 TYMFDWYGENIKGYSVQPGEIVAEGBEPTPEPVT-----ETPVDPTVTEEPVSE 599  
 QY 274 -----TPK----- 276  
 Db 600 LPDSYVIMELDKTKVKEGDVLIATIRVNNIKNLGAGYQIGIKYDKVLEAFNIETGDPIDE 659  
 QY 277 ----- 276  
 Db 660 GTWPAVGVTILKNRDLPTGVAINNVSKGIILFAAYVYFDDYEEGKSEDTGIIGNIGF 719  
 QY 277 -----EPAP--TTTKEPAPTTTTS 293







```

Db 764 ITSTISAPTTSTTTPQTSTSSPTTSTTTPQTSTSSPTTSTTSSPTTSTTSSPTTSTT 823
Qy 537 AAANTPKAPPTPKAPPTPKAPPTPKAPPTPKAPPTPKAPPTPKAPPTPKAPPTPK 592
Db 824 TSTQTSTISAPTTSTTSSPTTSTTSSPTTSTTSSPTTSTTSSPTTSTTSSPTT 883
Qy 593 -----KPAPELAPITTKPTSTTS----- 612
Db 884 TSSTTSGSGTTPSP---VPTTSTASVSKTSHVSVSKTHSQVTRDCHPRCTWTWFD 940
Qy 613 -DKPAP----- 617
Db 941 VDFPSPGHGDKETVNNIIRSGEKICRRPEEITRLQCRASHPEVSIHGLGVQVQCSRE 1000
Qy 618 -----TTPKG-----TAPTPK----- 629
Db 1001 EGLVCRNQDQGPCKCLNVEVRLCCETPKGCPVSTPVPAPSTPSGRATSTQSTSSW 1060
Qy 630 -----EPAPTPKPEAPPTPKGTPAPTTLKEPAPTTPKPKAPKEL 668
Db 1061 QKSRTTLVLTSTTSTPQTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTT 1119
Qy 669 APITTKGPTTSDKAPPTPKAPTTPKPEAPPTPKAPPTPKAPPTPKAPPTPKAPPT 727
Db 1120 --RTTSGPTTSTT--LAPITSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTT 1175
Qy 728 -TTKEPTTHKSDPELSPKALENSPKPCVPPTTKPAATKPEMITAKDKTT 786
Db 1176 STTSTPQTSKTSATSTTSGSGTTP-----SP-----VPTTSTTSTTSTTSTT 1226
Qy 787 ERDLRTPTTTAAPTAKPTKETAFTTBKTTESKITATTIQTSTTQ--DTTPPKITTLKT 844
Db 1227 -----SGPGTTPSPVSTSTTSAATSTTSTTSTTSTTSTTSTTSTTSTTSTTST 1281
Qy 845 TTLAPKVTTKKTIITTEINWKPEETAKPKORATNSKATPKPKPKAPKPK--TSTKPK 903
Db 1282 TS-APTST-----TSFGPTTPSPVPTTSTTSTTSTTSTTSTTSTTSTTSTT 1313
Qy 904 KTMPRVRKPKTTPTRKMTSTMPLNPTSRITAEAMLQTTT 943
Db 1314 -----TSFGSGTTPSPVPTTSTTSA--PTTSTTSTASTT 1346

RESULT 15
Q8IR51
AC Q8IR51; PRELIMINARY; PRT; 3432 AA.
DT 01-VAR-2003 (Tremblrel. 23, Created)
DT 01-VAR-2003 (Tremblrel. 23, Last sequence update)
DE CG32602-PB.
GN ORFNames=CG32602;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazei R.G., Champe M., Pfeiffer B.D.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu P.V., Berman B.P., Bayraktaroglu L., Beasley E.M.,
RA Bozkova D., Botchan M.R., Bouck J., Brokstein D., Brattier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.W., Cawley S., Dahlke C., Davenport L.B., Davies P.,

```

```

de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
Dodonson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
Dunbar K.J., Evangelista C.C., Ferraz C., Ferrer S., Fleischmann W.,
Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
Meunt S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
Nelson D.R., Nelson K., Nixon K., Nusskern D.R., Pacleb J.M.,
Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
Rainer K., Remington K., Saunders R.D., Scheeler F., Shen H.,
Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
"The genome sequence of Drosophila melanogaster."
Science 287:2185-2195 (2000).
[2]
SEQUENCE FROM N.A.
RP MEDLINE=22426065; PubMed=12537568;
RX Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverly T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodegren E.J.,
RA Swirskas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
"Finishing a whole-genome shotgun: release 3 of the Drosophila
melanogaster euchromatic genome sequence."
Genome Biol. 3:RESEARCH0079-RESEARCH0079 (2002).
[3]
SEQUENCE FROM N.A.
RP MEDLINE=22426070; PubMed=12537573;
RX Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Swirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
"Transposable elements of the Drosophila melanogaster euchromatin:
a genomics perspective."
Genome Biol. 3:RESEARCH0084-RESEARCH0084 (2002).
[4]
SEQUENCE FROM N.A.
RP MEDLINE=22426069; PubMed=12537572;
RX Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
"Annotation of the Drosophila melanogaster euchromatic genome: a
systematic review."
Genome Biol. 3:RESEARCH0083-RESEARCH0083 (2002).
[5]
SEQUENCE FROM N.A.
RP FLYBASE;
RL Submitted (SRP-2002) to the EMBL/GenBank/DBJ databases.
[6]
SEQUENCE FROM N.A.
RP FLYBASE;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003495; AA009587.1; -
DR FlyBase; FB00052602; CG32602.
SQ SEQUENCE 3432 AA; 349492 MW; FE907AC75578657A CRC64;

Query Match 20.1%; Score 1082.4; DB 2; Length 3432;

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 20.4912 Seconds  
(without alignments)  
4723.689 Million cell updates/sec

Title: SEQ1-E  
Perfect score: 5373  
Sequence: 1 MAWKTLPIYLLLLSVFVIQ.....DMDYLRVNPQGIINPMLS 1006

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.1

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: PIR 79:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1270.2	23.6	3020	2	A43932
2	1108.6	20.6	1664	2	T18262
3	1002.8	18.7	1274	2	T16251
4	990	18.4	1489	2	T31108
5	973.9	18.1	7962	2	I38346
6	969.9	18.1	2187	2	T30826
7	952.1	17.7	3570	2	T45025
8	892.3	16.6	1367	1	S48478
9	874.6	16.3	1188	2	S49915
10	845.4	15.7	6642	2	T29757
11	844.6	15.7	3507	2	T34513
12	844.3	15.7	1229	2	T25697
13	822.8	15.3	1151	2	T18535
14	822.5	15.3	1344	1	A35175
15	811.2	15.1	2897	2	B48666
16	811.2	15.1	3256	2	A48666
17	807.6	15.0	5762	2	A41819
18	757.4	14.1	1832	2	T31113
19	754.1	14.0	4135	2	T42629
20	750	14.0	2232	2	T34434
21	742.3	13.8	2142	2	B35098
22	736.8	13.7	3942	2	T42730
23	736.3	13.7	5262	2	T03454
24	735.6	13.7	761	2	C94672
25	735.4	13.7	990	2	I51618
26	733.5	13.7	924	2	S27923
27	729	13.6	971	2	T19431
28	728.5	13.6	4548	1	S00657
29	728	13.5	2225	2	T26063

## ALIGNMENTS

## RESULT 1

A43932

musin 2 precursor, intestinal - human (fragments)

N;Alternate names: musin SMUC-41

C;Species: Homo sapiens (man)

C;Date: 10-Mar-1993 #sequence revision 12-Apr-1996 #text change 09-Jul-2004

C;Accession: A49963; A45106; A43932; B33532; A61257; P00328; P00329

R;Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.

J. Biol. Chem. 269, 2440-2446, 1994

A;Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the

A;Reference number: A49963; MUID:94132002; PMID:8300571

A;Accession: A49963

A;Molecule type: mRNA

A;Residues: 1-639 <GU1>

A;Cross-references: UNIPROT:Q02817; UNIPROT:Q14884; GB:I21998

R;Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.

J. Biol. Chem. 267, 21375-21383, 1992

A;Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both upstr

A;Reference number: A45106; MUID:93016075; PMID:1400449

A;Accession: A45106

A;Status: not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 626-1895 <GU2>

A;Cross-references: GB:M94131; NID:G186395; PIDN:AAA59163.1; PID:G186396

A;Note: sequence extracted from NCBI backbone (NCBIP:116706)

A;Accession: B45106

A;Status: not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 2037-3020 <GU3>

A;Cross-references: GB:M94132; NID:G186397; PIDN:AAA59164.1; PID:G186398

A;Experimental source: colon

A;Note: sequence extracted from NCBI backbone (NCBIP:116698)

R;Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen, G.M.

J. Clin. Invest. 88, 1005-1013, 1991

A;Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polymorp

A;Reference number: A43932; MUID:91358717; PMID:1885763

A;Accession: A43932

A;Molecule type: DNA

A;Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>

A;Cross-references: GB:M74027; NID:G188863; PIDN:AAA59875.1; PID:G188864

A;Note: sequence inconsistent with the nucleotide translation

R;Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lampert, D.T.A.; Kim, Y.S.

J. Biol. Chem. 264, 6480-6487, 1989

A;Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and eviden

A;Reference number: A33532; MUID:89197956; PMID:2703501

A;Accession: B33532

A;Molecule type: mRNA

A;Residues: 1316-2193 <GU4>

A;Cross-references: GB:M22405; NID:G188873; PIDN:AAA36334.1; PID:G188874

A;Experimental source: intestine

R;Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.

hypothetical prote  
MHC class III hist  
Bassoon protein -  
MHC class III hist  
ascites sialoglyco  
hypothetical prote  
microtubule-associ  
hypothetical prote  
ALR protein - huma  
musin, tracheobron  
versican precursor  
hypothetical prote  
hypothetical prote  
probable tenascin  
BPLF1 protein - hu



Db 2282 VEIVKVECEPPMPPTCSNGLQVRVEDGCCWHMECCDCTGCGDPHYVTFDGLYSYQ 2341

Qy 968 GETPHMLLRHVFHFPEVTPDM-----EISPSVDNFVYIDNVCHCDNDKVCSPRTLIVRHETQEVLIKTVH 2395

Db 2342 GNCITYVLVE-----YLRFVPNOGIIN 1002

Qy 990 -----YLFVPNOGIIN 1002

Db 2396 MPMQOVQVNROAVALPKYKYLEVQSGINYVDIPELGVLS 2440

RESULT 2  
T18262

S-layer protein - Clostridium thermocellum  
C:Species: Clostridium thermocellum  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C/Accession: T18262  
R/Fujino, T.; Bequin, P.; Aubert, J.P.  
J. Bacteriol. 175, 1891-1899, 1993  
A>Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulose  
e.  
A/Reference number: Z18847; MUID:93209931; PMID:8458832  
A/Accession: T18262  
A>Status: Preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-1664 <FU>  
A/Cross-references: UNIPROT:Q06852; EMBL:X67506; NID:G296879; PID:G296881; PIDN:CAA47841

Query Match 20.6%; Score 1108.6; DB 2; Length 1664;  
Best Local Similarity 21.0%; Pred. No. 2.1e-15;  
Matches 359; Conservative 118; Mismatches 371; Indels 864; Gaps 69;

Qy 4 KTLPIYLLLLLLLVFV-----IQOVSSQE 26

Db 6 KVLSILLTLLIISTTSVMNSFAEATPSIEWLDKTEVHGVDVTATIKVNNIRKLGYQ 65

Qy 27 LSK-----GRCFSEFERREDCDAQCCKYDKCPDYBFCAEVKDNKKNR 73

Db 66 LNKFDPVELQVPDATGBEFT-----DKSMF-----VNR 95

Qy 74 ----TKKPTP-----KPVVDE----- 87

Db 96 VLTNSKYGPTPVAGNDIKSGINFATGVNNLTAYKSSGIDEHTGIIGCFGKVLKQNT 155

Qy 88 -----AGSGLDNG-----FKVTTPTSTTQHKNKVSPKITTAKPINPR 127

Db 156 SIRFEDTSLMPCGISGTSIFDWDAAITGYEVIQPLI-----VVEAPL--- 200

Qy 128 PSLPNSTSKETSUTVNKETT-----VE 151

Db 201 -----KDASVALELDKTKVKVGDIITATIKIENMKNFAGYQLNIKYDPTMLEIE 250

Qy 152 TKTTTTNKQT-----STDGKEKTS-----AKE 175

Db 251 LETGSATAKRTPWVGTVLQSDNYGKTAVANDVGAGHINFAYEAYSNLTKYRETGVAAE 310

Qy 176 TQSIEKTSAKDLAPTSKVLAKPTPKAETTT-----KGPALTTP--- 213

Db 311 TGLICKIGFRVLKAGSTAI-----RFEDTAMPAGIEGTMYMDWYGENIKGYSVOPGEI 365

Qy 214 -----KEPTPT-TYPEASTTPKEPTPT----- 235

Db 366 VAEGEEGERPEETEVPTEPVDPTPTVTEEPVPSLPDSYIMELDKTKVKVGDIITAT 425

Qy 236 -----TIKSAPT----- 242

Db 426 IKIENMKNFAGYQLNIKYDPTMLEIAELETSAGIAKRTPWVTGGTVLQSDNYGKTAVAN 485

Qy 243 ----- 242

Db 486 DVGAGIINFAYEAYSNLTKYRETGVAAETGIIKGIGFRVLKAGSTAIRFEDTAMPAGIEG 545

Qy 243 -----TPKEAPTPTTKPEAPTPTTKPEAPT- 273

Db 1534 RAEFATVVLHFLTKVKGQIMSKLATIDISNP 1565

RESULT 3

Tl6251

hypothetical protein F35A5.1 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004

C;Accession: Tl6251

R;Leimbach, D.

submitted to the EMBL Data Library, January 1996

A;Description: The sequence of C. elegans cosmid F35A5.

A;Reference number: Z18485

A;Accession: Tl6251

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-1274 <LET>

A;Cross-references: UNIPROT:Q20007; EMBL:U46675; NID:gl166613; PID:gl166621; PIDN:AAB526

A;Experimental source: strain Bristol N2; clone F35A5

C;Genetics:

A;Gene: CESP:F35A5.1

A;Map position: X

A;Introns: 1272/2

Query Match 18.7%; Score 1002.8; DB 2; Length 1274;

Best Local Similarity 23.1%; Pred. No. 2e-13;

Matches 330; Conservative 99; Mismatches 309; Indels 692; Gaps 75;

QY 74 TKKKPTP-----KPP--VDEAGSLDNGDFKVTTPDTSTTQHNVKSTSPKITTAKP 123

Db 2 SRAPPTPIKNPAKKWPWESVDE----- 25

QY 124 INRPRLPNNSDTSKETSIVANKETVETKTTTINKQTSIDGKEKTTSAKETQSIKTS 183

Db 26 -----EEMEVEDEETAPSK-----LEKKPSLKR----- 49

QY 184 AKDLATSKVLAKPTPKA-----ETTTKGPALTTTP-----KBPTP- 218

Db 50 -KD-APTQPV--PFGAPSVPIKNPVKKWKAPWEDDEMEAPAAPVPAKKVRDPSK 104

QY 219 TTPKEASTTPKE-----PTPT-----TTKSAPTTP 244

Db 105 KVPKAPRDSQPKIMAAKKEBELPAVPTPVKNPVKKFKAPWEDDEVVDVKDAFTVP 164

QY 245 -----KEPA-----PTTKSAPTTPKSP--APTTPKEPA-----PTT 274

Db 165 AKKTPVLKKKEPAAAKPRDPSPKAAPSKHEHPDIVPPTPIKNPAKKWKPPWEDDEVPT 224

QY 275 P-KEPAPTTPKEPA-----PTTKSAPT----- 296

Db 225 EIKSEPEATRKKVPALKKKKSPSTSVKVSDESPSTKKVPVKKKEPEVPTPIKNPTKKWKPPW 284

QY 297 --TP-----KBAP-----TTPKKPAPTTPKEPAPTTPK 323

Db 285 EDETVEVEKPPVPEKAPVLKKDPAPAAKARDPSKAAAPKKVEPSSPVVP-PPEVK 343

QY 324 EPT-----PTTPKEPAPTTPK-----EPAPTTP-KEPAP----- 350

Db 344 NFVKYKPPWVDDPEAEVKKPSAPEKTPVLKRKEPEPSSSTTPSSDPSPKAAAPAVKP 403

QY 351 --TAPKKPAP-----PTTKKEP-----TTPKEPAPTTP-- 369

Db 404 RDSQPKATPLQADPKAQEVPTPVKNPVKKYPWEDDEDVVEVKQAPAPAKKTPVL 463

QY 370 --KEPA-----PTTKKEP-----GPTTP-KEPA----- 389

Db 464 KRKEPAKQDAPKATSKTPTPEIKKDPVKPRDSGPKVAAKPDGAQADATPVKNPVKKWR 523

QY 390 -----PTTKKSAPT--TTKEPAP-----TTTKSAPTTPKEPSPTTTKE 425

Db 524 PPWEDDETADDVSKPTDAKKTTPSLAKKDPAPAKESLKPADTK-APAKPRDPS--KK 579

QY 426 PAPTTPKEPAPTTPKK-----PAPTTPKEPAPTTPKEP-----APTTPKK 465

Db 580 VAPTAPEKKTVPVLAKKEPAGPADSKTKPEKSKPRDPSFKKAVPAKVPKTEVAPAAVKK 639

QY 466 PAP-----TAPKEPAPTTPKETAPTTP-----KKLTP-----TTPEKLAP 500

Db 640 PEPISKPDXTAPKKAEPNSP--VVVPTPVKNPVKKWKPPWEDDDAPAKPVSLPEPEKKTTP 697

QY 501 TTPEKAPATTPE-----ELAPTTPPEPTTP--TTPEEPAPTTP----- 534

Db 638 VLAKK-APTKEPDEAAADPVSGPSSKDKLAKKAFVKPRDPSPMKAVPIKPAKTEVPEA 756

QY 535 -----PK--AAAPNTPKEPAPTTPKEP----- 554

Db 757 VVKKEPVPVAKSRDPSPKKAKAEPNSPVVP-PTPVKNPVKKWKPPWEDDDAPAEPVNVVPP 815

QY 555 -----APTTPKEPAPTTPKETAPTTPKTGTAATTLKEPAPTTP-PKKPAPKE----- 598

Db 816 EKKTPVLAKKTVPKPRDPSPKKAVPAKPFSTKTADPVSVKKPEPVSKPEPKKAEPN 875

QY 599 --LAPTTPKEPT-----STTSKAPAPTTPKTGAPT--TPKEP-----APTTPKEP 639

Db 876 FVVPTPVKNPVKKWKPPWEDDDDETEEVKPSSEPEKKTVPVLAKKEPEKPKDAPKAAKP 935

QY 640 APTTPKTGTAATTLKEPAPTTP-----TPKKAP-----KELAPTTPKTGPTS-----TTS 681

Db 936 RDPSPKAVPE--KEPAKVAAPKPRDLSPKKAIPANTQEAAPTTPVKNPVKKWKPPWEDD 993

QY 682 DKP-----APTTPKET-----APTTPKEPAP-----TTPKKPAPTTPPETPTTSEVSTP 726

Db 994 DEPAEVSAPPEKKTVPVLAKKAPAKPRDPSFKKAAAPVAAKPDPKIPEVP-----P 1044

QY 727 TTTKEPTTIHKSP-----DESTPELSA-BP--TPKALENSPKPEGV-----P 765

Db 1045 TPVKNPVKKWKPPWEDDDDESPVSAPEPEKKTVPVLAKKAPTTPATKPKDSEAAADPVSGP 1104

QY 766 TTKTPAAT-----KPEMTTTAKDKTTERDLRTTPETT--TAAPKMTKETATTTTEKTES 817

Db 1105 TSKDPKLSKKAPEKPEKPTTDPKDDKLKPSPAKKKEPAPEAPAAKPKWKPPWDDDDPEPEA 1164

QY 818 KITATTITQVTSITTTQDTPPKITTLTTLTTLAPKVTTTKITTTTEIMNKPEETAAPKDDRA 877

Db 1165 DFTVPAPSKKPDTEADPLG-----GPKTKDPK-----LNKKAPAEKPTPK- 1206

QY 878 TNSKATTPKPKQPKTKAPKPTSTKPKTMPRVRKPK-----TTPTPRKMTSTM 925

Db 1207 -----PKPEKYSKEPPEKTEPPKP-AAPEKKWKPPWEDDDPEADFTMPAPKK----- 1253

QY 926 FELNFTSRIAEAMLQTTTRPNQTPNSKLVVNPKSED-----AGGAGGETP 971

Db 1254 -----PDTEPDADSLGGPKPKDP 1271

RESULT 4

T31108

cyst germination specific acidic repeat protein precursor - Phytophthora infestans

C;Species: Phytophthora infestans (potato late blight agent)

C;Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004

C;Accession: T31108

R;Goernhardt, B.

submitted to the EMBL Data Library, April 1998

A;Reference number: Z20986

A;Accession: T31108

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-1489 <GOE>

A;Cross-references: UNIPROT:O96449; EMBL:AF061185; NID:g3851513; PID:g3851514; PIDN:AAC

C;Genetics:

A;Gene: car90

Query Match 18.4%; Score 990; DB 2; Length 1489;

Best Local Similarity 26.1%; Pred. No. 4.3e-13;

Matches 391; Conservative 53; Mismatches 415; Indels 640; Gaps 70;

```

QY 27 LSCKGRCEPESPERGEC-----DCDAQCKKYD--KCCPDYESFCAEVKON----- 69
Db 43 IPCSGVGAEPV--GTACPKAGDVATSDCPYLLSYNGAVCVAPVDAECALHDDMWGCEP 100
QY 70 ----- 69
Db 101 PKGYTSAVEAETIAAANGBSSGNGTGHDEVVQVGDDEEEIIPARVNYDTTVDTPIGNVCE 160
QY 70 --KXNRTKKKPTKPPVVVDVAGSLDNGDF-----KV----- 99
Db 161 VATETATQGHATEGGKYDTPSTQVQDGYNTIHYGSTTTEGVTKGGYGPTDAKVIDGE 220
QY 100 -----TPDTSST-----QHNKVS 113
Db 221 TYLDYPTGITEIIBDGTTPGGYGTDTGGTTGGYTTVDNTHETGAGGYDACTREYS 280
QY 114 TSPKI-----TTAKPIN-----PPSLP- 131
Db 281 TTPVGVSTTEGQHWGTVGYPSEDETEAPTEGTYVVPREETTAAPSEDTTYAPREVTPY 340
QY 132 -----PNSDT--- 136
Db 341 APTEKPYDVBEITYVTEESTYAPTKSETNAPTERMHYAHIKPCDTEVTWYAPTEETYA 400
QY 137 -----SKETSLVYNKETTVE--TKETT--TTNKQTSDDGKBKTTSAKETQSIKTSAK 185
Db 401 PTEETTYAPTEETTYAPTEETPYEPTTEETTYPTTEETTYAPTEET-- 446
QY 186 DIAPTSKVLAKPTKPAETTTKGALTTPKPEPTTT----- 220
Db 447 -YAPTEKTYAPT--BEITYAPTEETPYEPTTEETTYAPTEETPYEPTTEETTYAPTEET 502
QY 221 --PKE-----PASTPKPEPTTTIKSAPTTKPEPAPT--TTKSAPTTKPEPAPT--TTKEP 270
Db 503 YAPTEETTYAPAEETPYEPTTEET--TYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 561
QY 271 APPTKPEAPTTKPEAPTTKSAPT-----TPKPEAPTTKPEAPTTKPEAPTTKPEAPTT 320
Db 562 ABETPYEPTTEETTYAPTEET--YAPTEETWYAPIETTYAPTEETTYAPAEETPYEPTTEE 620
QY 321 TPKEPPTTKPEP-----AP--TTKEPAPTTPKEP-----APTAPKPA 357
Db 621 TYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYA 680
QY 358 PTPKPEAPTTKPEPAPT--TTKPSPTTKPEP-----APT--TTKSAPT----- 398
Db 681 PTEETTYAPTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEE 740
QY 399 TTKEP-----APT--TTKSAPT--TPKPSPTTKPEP-----APTTPKPEAPTTKPA 443
Db 741 TYGTEETTYAPAEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 800
QY 444 PTPKPEAPT--TPKPEAPTTKPEAPTAPEK-----APTTPKETAPTTP 487
Db 801 PTEETTYAPTEETPYEPTTEETTYAPTEETPYEPTTEETTYPTTEETTYAPTEETTYAPTEK 860
QY 488 KKLTPTEKELAPT--TPEKP-----APTPTEELAPT----- 517
Db 861 TYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 920
QY 518 ---TPREPTTTPEEP-----APTPKAAAPNTPKPEAPT--TPKPEAPTTPKEP--- 562
Db 921 PAETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTTEE 980
QY 563 ---APTTPKETAPTTPKGTAP-----TTLKEPAPTTPKPP-----APKE---LA 600
Db 981 TYAPTEETTYAPTEETWYAPIETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYA 1040
QY 601 PT-----TTKEPTSTTSKPA-----PTTPKGTAPTTPKPEAPTTKPEAPTTTP 644
Db 1041 PTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEE 1100
QY 645 KGTAFT--TLKEPAPTTPKPP-----APKE---LAPT-----TTKGPTSTTSK 683

```

## RESULT 5

I38346

elastic titin - human (fragment)

C:Species: Homo sapiens (man)

C&gt;Date: 29-May-1998 #sequence revision 29-May-1998 #text\_change 09-Jul-2004

C:Accession: I38346

R:Labelit, S.; Kolmerer, B.

Science 270, 293-296, 1995

A:Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.

A:Reference number: A57430; MUID:96026330; PMID:7569978

A:Accession: I38346

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-7962 &lt;RES&gt;

A:Cross-references: UNIPROT:Q10465; EMBL:X90569; NID:gl017426; PIDN:CAA62189.1; PID:gl01

C:Genetics:

A:Gene: GDB:TTN

A:Cross-references: GDB:127867; OMIM:188840

A:Map position: 2q31-2q31

## Query Match

Best Local Similarity 18.1%; Score 973.9; DB 2; Length 7962;

Matches 390; Conservative 107; Mismatches 400; Indels 2121; Gaps 92;

```

QY 5 TLPIYLLLLSVFVIQVVS---SOELSKGR-----CF 34
Db 4844 TQPI-----KVSNAKDSREIRSGKQYLSYLSNAHLTVLKVKGDSQGYTCY 4891
QY 35 ESFERGRCDDAOCKYDKCCPDYESFCAEVKON----- 69
Db 4892 AVNEVGKD-SCTAQLNIKERLIP--PSFTKLSEIVETEGNSFKLEGRVAGSQDITVAV 4948
QY 70 KKNRTKKKPT-----P 80
Db 4949 YKNNIEIQTNSCBITFKNNLTVLQVRKAGMNDAGLYCTKVSNDAGSALCTSSIVIKEPK 5008
QY 81 KPPVVDE-----AG-----SG----- 91
Db 5009 KPPVFDQHLTPVTVEGEYVQLSCHVGSSEPIRQWLKAGREIKPSDRCSFSFASGAVL 5068
QY 92 -----LDNGDF-----KVTTPDTSHTQNKVSTSPKITTAKPINRPSLPNSD----- 135
Db 5069 ELRDVAKADSGDYVCKASNVAGSDT-----TKSKVT-----IKDKPAVAPATKAAV 5115
QY 136 -----TSKETSLSLVNKETT----- 149

```

Db 5116 DGRFFVSEPOSIRVVKTTATFIAGVGDPINPVKWKQWRQNLQNGRQVFIHOKGDEA 5175  
Qy 150 -VETKETTTN-----KQTSIDG----- 166  
Db 5176 KLEIRDTTKDGLRYCAVAFNEHGEIESNVNLQVDERKKQEKIEGDLRLAMLKKTPIILKKG 5235  
Qy 167 -----KETTSAKETQSIE 180  
Db 5236 AGEBEEDIMELLKNVDPKPEYKARMYGITDFRGLLOAFELLKQSQEBETHRLTEIEIE 5295  
Qy 181 K-----TSAKD----- 186  
Db 5296 RSEDERKEPEELVSEFIQRLSOTPEVTLIKDINENOTVLKNDANVFEIDIKINPEILKLSW 5355  
Qy 187 -----LAPTSK----- 192  
Db 5356 YKGTEKLEPSDKFEISIDGRHTLRVKNQLKQDNVRLVCGPHIASAKLTVIEPAWHRH 5415  
Qy 193 ----- 192  
Db 5416 LQDVTLKEGQTCMTVQFSPNVKSEWFRNGRILKPGRHKTEVEHKVHLTIADVRAED 5475  
Qy 193 ----- 192  
Db 5476 QGQYTCYEDLETSABELRIEAEPIQTKRIQNIWSEHQSATPECEVSFDDALVWYKGP 5535  
Qy 193 -----VLAKPTPKAETTTKGPAITPKPE--- 215  
Db 5536 TELTESQKYNFRNDGRCHYMTIHNVTDDGVSIVARLEPRGEARSTABLYLTKEIKL 5595  
Qy 216 -----TPPT-----TPKE----- 223  
Db 5596 ELKPPDIPDSRVEIPTMPIRAVPEEIPPVAVPEVFLLLPTPEKKPPPKRIBVTKKAVK 5655  
Qy 224 -----PASTPKPE-----PTPTTI----- 237  
Db 5656 KDAKKVVAKPKEMTPREIIVKKPPPTTLIPAKAPEIIDVSSKABEVKIMTIIRKKEVQK 5715  
Qy 238 ----- 237  
Db 5716 EKEAVYKQAVHKEXEVFIESPEEPYDELEVETPTPEPQYVEEPEDEYERIKVEAKK 5775  
Qy 238 ----- 237  
Db 5776 EVHEEEDPEEQEYVERBEGYDEGBEWEAEAYQEREVIQVQKEVYEEESHKRVKAPVP 5835  
Qy 238 -KSAPITPK----- 245  
Db 5836 EKKAPPPPKVKKPVIEKIEKTSRMEBEKVQTKVPEVSKKIVQPKSRTPVQEEVIEV 5895  
Qy 246 -----EPAPITTK 253  
Db 5896 KVPVAVTKKMWISEEKMFASHTEEYSVTVPVQKGIIVTBEKIHVAVSKRVEPPP----- 5951  
Qy 254 SAPITPKPA-----PTTKKEPAPTPEPAPTTPKEPAPTTPKSAPTTKPA----- 302  
Db 5952 KVPBLPKPAPEEAVAPVPIPKKVPPEAPKVEVP-----KKPVPEEKPPVPVPKKEPAAP 6006  
Qy 303 -----PTTPKKAP-----TPPKPE-----PAPTTPK 323  
Db 6007 PKVEVPKKPVPEEKIPVPAVAKKEAPPAKVPEVQKGVVTEEKITIVTQRESEPPPAVPE 6066  
Qy 324 RPTPTPKPE-----PAPTTPKEPAPTTPKEPAPTAPKAPAP----- 358  
Db 6067 IPKKVPEERKVPKKEEVPVPPKVPKPA--LPKKVPEEKVAVPVPAVAKKAPPPPAEVS 6124  
Qy 359 ----- 358  
Db 6125 KTVVEKRFVABEKLSPAVQORVEVTRHVSABEWSYSEBEEGVSISVYREEEEBEEEE 6184  
Qy 359 -----TPPKPA----- 365

Db 6185 AEVTEYVMEBPPEEYVVEEKLHIISKRVEAPAEVTERQEKKIVLKPILPAKIEBPPPAK 6244  
Qy 366 -PTTPKE-----PAPTTPKE--PSPTTPKEP-APTTPKGAPTTT--KEPAP----- 405  
Db 6245 VPEAPKIVPEKKVDPAPVPKKEKVPKVPKVPPEPKVPPEKVPKVIKMEHELPAKVTEK 6304  
Qy 406 -----TTTKGAPT--TPKEPSPTTTKE-----PAPT----- 430  
Db 6305 HMOITQEEKVLVAVTKKEAPPKARVPEBPFGAVPEEKVLKLPKREBPAPKVTFRKRV 6364  
Qy 431 -----PK-BAP----- 436  
Db 6365 VKEEKVSIAPKREPQPIKEVTIMBEKRAYTLEBEAVSQREBEYEEYEDYKEFEY 6424  
Qy 437 -----TPPKKAPT--TPKEPAPTTPKE----- 457  
Db 6425 EPTTEYDQVEEYEEERYEYEBHEYITEPEKIPVVKVPPEBPVTKKAPPAKVLKXAV 6484  
Qy 458 -----PAPTTPTKPAPTAPKEP----- 474  
Db 6485 PSEKVPVPIPKKLKPPPKVPPEEKVPPEEKIHSITKREKEQVTEPAAKVPMPKRVVA 6544  
Qy 475 -----APTTPKETAP----- 487  
Db 6545 EEKVPVPRKEVAPPVVRVPEVPKLEBEVAFBEVVTHVEEYLVVEEYIHEEBSFITE 6604  
Qy 488 KKLTPITTPPEKLAPTTPKEP-----APTTPPELAPT----- 517  
Db 6605 EEWVPIPVKV--PEVRKVPPEKVPKVPKKEAPPKAPKAPVPEVPKPEEKVPLIPKKEK 6663  
Qy 518 -----TPBEFTPTTPPEAPAP-----TPKAAAPNTPKPEPAPTTPKEPAPT-----TPK 560  
Db 6664 PPKAKVEVPKVPPEKVPVVPKKEVAPPKAPKVPPEKVPKVP--EKKVPVPAPKKEVAPP 6722  
Qy 561 EPAPTTPKETAPTTPKGT-APTTLKEPAPTTPK----- 593  
Db 6723 AKVPEVPKKLIPBEKPTTPVKKVEAPPKVPKKEPFPVVPVVALPOEEVLPFEEIPEE 6782  
Qy 594 ----- 593  
Db 6783 EVLPSEEEVLPSEEEVLPSEEEVLPSEEEVLPSEEEVLPSEEEVLPSEEEVLPSEV 6842  
Qy 594 -----PAP-----KELAP-----TT 603  
Db 6843 KPKVPVAPVPEIKKVKTERKKVWIPKKEEAPPKAPKVPPEKKEKRIILPKSEEVLPSEV 6902  
Qy 604 TKEP-----TSITSDKEPAPTTPKGTAPT----- 627  
Db 6903 TERPEEPISEEEIPEEPPPSIEVEEVAPPRVPEVIKKAPEAPTTPVKKEAPPAKUSK 6962  
Qy 628 --PKE--PAPTTPKEPAP----- 641  
Db 6963 KIPKEKVPVQKKEAPPAKVPPEKVPKVPKKEVLPKKEVAPPKGRVLEEKVSAPRQ 7022  
Qy 642 ----- 641  
Db 7023 EVVVKERLELEVWAEVEEIPPEEPEHEVEEYEEGEFHEVEEFIKLEQHREVEHEHREK 7082  
Qy 642 -----TPPKGTA-----PPTL--KEP--APTTPK-- 662  
Db 7083 VHRVIEWFAEVEEVEFKKAPKAPKGEISEKIIIPKPPPTKVPKPPKAPKVPPEVKIV 7142  
Qy 663 -----PAPKELAPTTPKGTSTTSKPAPTTPKETAP-----TTPEPAPTTPK 706  
Db 7143 VEKVRVPEEPRVPTKV-----PEVLPKPEVVEKVPPEKPAKKEAPKVPPE 7192  
Qy 707 KPAPTPE-----TPPTTSEVSTTTTKE-----PTTIHKSPDESTBELSAP- 750  
Db 7193 APKEVVEKVPVPPPKPEV--PPTKVPEVPEKAAVPEKKVPEALPPKPESTPPEFEFE 7251  
Qy 751 -TPKALENSPKPEGVPTT-----KTPAA--TKPEMTTTAKDKTTERDLRTP 794  
Db 7252 ESPSA---PPKKPEVFPVRVPEVPKVEVVPKVPKPAAPPKPEV-----TP 7293



```

QY 795 ETTAAAPKMTKATTTTEKTTESKITATTQVTSITTTQDITTEPKITTLTKTTLAPKVTTT 854
Db      |||      |||      |||      |||      |||      |||      |||      |||
7294 VKVPEAPK-----EWPEKKV-----FVPPPKKEVPPTKVPEV-----PKVAVP 7333
QY      |||      |||      |||      |||      |||      |||      |||      |||
855 KKTITTT-----EIMNKPETA----- 871
Db      |||      |||      |||      |||      |||      |||      |||      |||
7334 EKVPEAIPPKPEPPPEVPEEPEVALEPPAEVVERPEPAAPPQVTVPPKNVPVEKKA 7393
QY      |||      |||      |||      |||      |||      |||      |||      |||
872 ----- 903
Db      |||      |||      |||      |||      |||      |||      |||      |||
7394 PAVVAKKPELPPVKVPEVPEVPEKKVPLVVPKKPEAPPAKVPEVPEVPEKKVAVPK 7453
QY      |||      |||      |||      |||      |||      |||      |||      |||
887 -----POKPTKAPKKPTSTTKP----- 903
Db      |||      |||      |||      |||      |||      |||      |||      |||
7454 KPEVPAKVPEVPPKPVLEEKPAVVPVERAABSPPEVVEEPEETAPBEEETAPBEKVPV 7513
QY      |||      |||      |||      |||      |||      |||      |||      |||
904 -----KTMERVRKPKTTP----- 916
Db      |||      |||      |||      |||      |||      |||      |||      |||
7514 AEEEEVPPPAVPEPKKIPEKKVPIKKPEAPPPEPEKPKLKPBPVPPPP 7573
QY      |||      |||      |||      |||      |||      |||      |||      |||
917 -----TPRKMTS----- 923
Db      |||      |||      |||      |||      |||      |||      |||      |||
7574 APBKEDVKEKIFOLKALPKKVPENPOVPEKVELTPLKVPGGKKVKLLPERKPEKEE 7633
QY      |||      |||      |||      |||      |||      |||      |||      |||
924 -----TMPELNPTSRIAEAMLQTTTR-----PNQ 947
Db      |||      |||      |||      |||      |||      |||      |||      |||
7634 VVLKSVLRKRBEERKVEPKLEKVKKPAVPPPPPKPVVEEVVPTVKREKIDPEPK 7693
QY      |||      |||      |||      |||      |||      |||      |||      |||
948 TPNSK-----LVEVNPKSDAGAGETPHMLLRHVFVPEVTP----- 986
Db      |||      |||      |||      |||      |||      |||      |||      |||
7694 VPEIKPAIPLPAPEPKPE-----AEVKT-----IKPPVPEPPTPIAAPVTVPVVGKAE 7745
QY      |||      |||      |||      |||      |||      |||      |||      |||
987 -----DMDYLPVRVNOGI 999
Db      |||      |||      |||      |||      |||      |||      |||      |||
7746 AKAPKEAAKPGPIKGV 7763

RESULT 6
T30826
nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse
N:Alternate names: alpha-NAC protein
C:Species: Mus musculus (house mouse)
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30826
R:Yotcov, W.V.; St-Arnaud, R.
Genes Dev. 10, 1763-1772, 1996
A:Title: Differential splicing-in of a proline-rich exon converts alphaNAC into a muscle
A:Reference number: Z20889; MUID:96312450; PMID:8698236
A:Accession: T30826
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2187 <YOT>
A:Cross-references: UNIPROT:P70670; EMBL:U48363; NID:g1666688; PID:g1666689; PIDN:AA8187
A:Gene: Naca
A:Map position: 10
A:Iron: 24/1; 1996/1; 2024/3; 2050/3; 2099/3; 2142/3; 2183/3
A>Note: differential splicing converts alphaNAC into a tissue-specific DNA-binding activ
C:Keywords: alternative splicing; DNA binding; transcription factor

Query Match 18.1%; Score 969.9; DB 2; Length 2187;
Best Local Similarity 21.7%; Pred. No. 1.8e-12;
Matches 351; Conservative 134; Mismatches 412; Indels 721; Gaps 67;

QY 4 KTLPIYLL-----LISVFIQVQVSGQELSCKRCFESPERG 40
Db      |||      |||      |||      |||      |||      |||      |||      |||
479 KNLPISALVNGAVSPVPAQAGLPTKDTTLOPLAIKESFSSQAS-----SLEVL 531
QY      |||      |||      |||      |||      |||      |||      |||      |||
41 RECDCAQCK-----YDKC----- 55
Db      |||      |||      |||      |||      |||      |||      |||      |||
532 SE---DVTYKTTGGPVPVVPATAGVATTTSRADSPAVIRADSCVSPNTVQPLKRS 588

```

```

QY 56 -----CPDYBSFCAEVKDNKKNRT-----KKK 77
Db      |||      |||      |||      |||      |||      |||      |||      |||
589 VTDPAAPRTAKTAPSTTSPVLASBGCPCVASSMALSPOQASVSETALALSPIPKSV 648
QY      |||      |||      |||      |||      |||      |||      |||      |||
78 PTPKPPV-----VDEAGSLDNGDFKVTTPTDSTTOH----- 109
Db      |||      |||      |||      |||      |||      |||      |||      |||
649 PFPDPPLAEISFSNARKVDVASHMESGSGRQGHDPASVTAKGTWVCLADSLDTSVSAS 708
QY      |||      |||      |||      |||      |||      |||      |||      |||
110 -----NKVSTSPKITTAKPINPRPSLPNSD 135
Db      |||      |||      |||      |||      |||      |||      |||      |||
709 KGSALSGASSPLYLEVFLPEAGLAVQGPGLKSLKLSPTPPSSKGPV--PSTGAPP-- 764
QY      |||      |||      |||      |||      |||      |||      |||      |||
136 TSKEISITVAKETIVETKE-----KDLAPTSKVLAKP-----TTTIN 159
Db      |||      |||      |||      |||      |||      |||      |||      |||
765 -SPGAPIVPTSESISSKQVPAEILPSPQKTPVETASRLISAVQSPKVDPIMSDVTISP 823
QY      |||      |||      |||      |||      |||      |||      |||      |||
160 KOTSTDGKEKTTSAKETQSIETSA-----KDLAPTSKVLAKP----- 197
Db      |||      |||      |||      |||      |||      |||      |||      |||
824 KKTISATAVPKOTSA--TLSLSKVPATVLSPPKAPVAPSNEATIVPTETPLSLKNALAAA 881
QY      |||      |||      |||      |||      |||      |||      |||      |||
198 -----TPKAEITTKGPALTT-----PKE----- 215
Db      |||      |||      |||      |||      |||      |||      |||      |||
882 TPKETLATSIPKVTSPSPQKTPKSVLSKGPAMTSSKATEIAASKVDVPSQFPKEVPLLQ 941
QY      |||      |||      |||      |||      |||      |||      |||      |||
216 --PTPTPKPEASTT-----PKEPTPTTIKSAPTTPKE--PAPTITTKSAPTTPK-- 260
Db      |||      |||      |||      |||      |||      |||      |||      |||
942 HVPTSPPKSPVSDTLGALTSPPPKGP--PATLAETPTVPKSPKPAASKKTATPESPEG 1000
QY      |||      |||      |||      |||      |||      |||      |||      |||
261 -----EPAPTITTKPEAPTTPKEAPTTPKEAP-----T 289
Db      |||      |||      |||      |||      |||      |||      |||      |||
1001 VTAVPLEIPECSKAPKATAAPKESATSSSKRAPKTAVSKBPSKGVTAVPLEISLPKE 1060
QY      |||      |||      |||      |||      |||      |||      |||      |||
290 TTKSAPTTPKEAPTTPKE--PAPTTPKE-----PAPTTPKEPTP----- 327
Db      |||      |||      |||      |||      |||      |||      |||      |||
1061 TSKSA--TPGEKSASSPKSPKTAGKETPPGGVTAVPPEISLPPKRETFONATPNSLAA 1118
QY      |||      |||      |||      |||      |||      |||      |||      |||
328 -----TTPKE-----PAPTTPKE-----APTTP 345
Db      |||      |||      |||      |||      |||      |||      |||      |||
1119 SSQKSGPKTSVPKETPPGGVTAMPLEIPSAQKAPKTAVPKQIPTPEDAVTILAGSLSP 1178
QY      |||      |||      |||      |||      |||      |||      |||      |||
346 KEPAITAPKPKAPTTP-----PAPTTPKEAPTTPKEAPTTPKSAPTT--PK 416
Db      |||      |||      |||      |||      |||      |||      |||      |||
1179 KAASKTAAPKEAPATPSPGVIAVSGEISPSPKKTSKTAAPKENSATLPKRSFKTAAPKE 1238
QY      |||      |||      |||      |||      |||      |||      |||      |||
372 PAPTTPKE-----PSPTTPKE--PAPTITTKSAPTTTPKEAPTTPKSAPTT--PK 416
Db      |||      |||      |||      |||      |||      |||      |||      |||
1239 -TPATSSSGVTAVPSEISPSPTTPASKGVVTLTPKGAPNALAE--SPASPKVKVKTAAPE 1296
QY      |||      |||      |||      |||      |||      |||      |||      |||
417 EPS--PTTTPKEAPTTPKEAPTTPKKAAPT-----PKEAPTTPKEAPTTPKTP 466
Db      |||      |||      |||      |||      |||      |||      |||      |||
1297 ETSITTPSQKIPKVAGPKESATPSPKKTPTKTAVPKETSAPSEGVTAVPLEIIPPSPKAP 1356
QY      |||      |||      |||      |||      |||      |||      |||      |||
467 APTAPKE--PAPTTPKETAPTTPKLTPTTPKLAAPTTPKEAPTTPPELAPTTPPEPTPT 525
Db      |||      |||      |||      |||      |||      |||      |||      |||
1357 KTAAPKETAPS--PEGATTAPVQI--PPSPKSGKAGSKETPTTP-----SPEGVTAA 1407
QY      |||      |||      |||      |||      |||      |||      |||      |||
526 TPBEP--APTTPKAAAP-----NTPKEAPTTPKEAPTTPKEAPTTPKEAP 564
Db      |||      |||      |||      |||      |||      |||      |||      |||
1408 PLEIPISSKKTSMKASPKETLTVTSSKKLSQTVGPKETSLGATAVPLEIPSHKKAAPT 1467
QY      |||      |||      |||      |||      |||      |||      |||      |||
565 TTPKETAPTTPKGAPTTPKKEAPTTPKKAAPKELAP-----TTTKEPTSTTSKPAPTTP 620
Db      |||      |||      |||      |||      |||      |||      |||      |||
1468 VDPKQVPLTPSKDAPTTLAE--SPSPK--APKTAAPPSESVTTVPP-----EKPA-TPQ 1519
QY      |||      |||      |||      |||      |||      |||      |||      |||
621 KGTAAPTTPKEAPTTPKEA-----PTTPKGTAAPTTPKEAPTTPKKAAPKELAPT 671
Db      |||      |||      |||      |||      |||      |||      |||      |||
1520 KASGTTASKVPVPAETQEVAVSSRETPTVPAVPVKNPSPSSHKTSKTELKEADATLPPS 1579
QY      |||      |||      |||      |||      |||      |||      |||      |||
672 TTGPTSTTSKPAPTTPKETAPTTPKKEAPTTPKKEAPTTPPETPTTSEVSTPTTKE 731
Db      |||      |||      |||      |||      |||      |||      |||      |||
1580 PTKSPKIPSSKK--APRTSAP-----KEFPASPSIKPVTTSLAQT-----A 1618
QY      |||      |||      |||      |||      |||      |||      |||      |||
732 PTHHSPDESTEP--LSAETPKALENSPKFPGVPTTKTPAATKPEMTTTAKDKTTERD 789

```

```

Db      1619 PPSLOKABSTIPKENLAA---PAVLPSVSSKSPAAP-APASASLSP---ATAAPOTAPKE 1671
Qy      790 LRTTP-----ETTTA-----APKMTKETATTTEKSKITATITTTQVSTTTQ 832
Db      1672 ATTIPSCKKAATEPIETSTAPSLEGAPKETSEIV-----SKVLSPPPKKASSSK 1724
Qy      833 DTTPPKITTL-----KTTTLAPKVVTTTKKIIITTEIMNKPEETAKPKDRATNSKATTPKPQ 888
Db      1725 RASTLPAILPLPSLKASVLSPTATSSGK-----DSHISPVSDACSTGTTT--PQ 1771
Qy      889 KTKAPKK-----PSTNKKPKT 905
Db      1772 ASEKLPKSGKPTAFTEMLAAPAPESALAIATPIQSPGANSNSASSPKCPDPSSKKDTKG 1831
Qy      906 MPR--VRKEKTTPTPRKMTSTWPELNPTSRISAEAMLQITTRPNQ----- 947
Db      1832 LPSAVALAPQTVPEK-----DTSKAIETLLVSPAKGSDCLHSPKGPVGSQVATP 1881
Qy      948 -----TPNSKLVEVNPKSEDAGAGETPHMLL 975
Db      1882 LAAFTSDKVPPEAVSASVAPKAPAPASLTLAPSPVAPLPKQPKLLESAPGSVLESFSL- 1940
Qy      976 RDHFVMEVPTPDMYLPV-----PNQGIILN-----PML 1005
Db      1941 -----PVPABEDELPLIPPEAVSGGEFFQPILVNMPAPKPAAGTAPAPASAKQFVL 1991

RESULT 7
T45025
Mucin MUC5B, tracheobronchial [imported] - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T45025
R;Desseyn, J.L.; Guyonnet-Duperat, V.; Porchet, N.; Aubert, J.P.; Laine, A.
J. Biol. Chem. 272, 3168-3178, 1997
A:Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alternat
A:Reference number: Z22899; MUID: 97166151; PMID: 9013550
A:Accession: T45025
A:Status: preliminary; translated from GH/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3570 <DES>
A:Cross-references: EMBL:Z72496; MID:g1834502; PIDN:CAA96577.1; PID:g1834503
A:Experimental source: placenta
C:Genetics:
A:Gene: MUC5B

Query Match 17.7%; Score 952.1; DB 2; Length 3570;
Best Local Similarity 13.0%; Pred. No. 7.8e-12;
Matches 426; Conservative 116; Mismatches 412; Indels 2319; Gaps 96;

Qy      6 LPYLLLLLVFVIQVSSQBLSCK-----GRCFESFER----- 39
Db      3 LPV-----STVCREV-----CRMSSWYNGHRPEPLGGGDPETENLRQRGVQVCPV 50
Qy      40 -----GRECD-----AQQKYDKCCDYE--SFCAEVKONKK 71
Db      51 LADIECAAQLPDMPLBELGQVQVCDRMRGLMCANSQQSPPLCHDYELRVLCCEY----- 105
Qy      72 NRTKKKPKP-----PVDVAGGLDN----- 82
Db      106 --VPCGSPAPGTSPQPSLSASTEPVPTQTATATEKTLTWTPSIRSTAALTSQTGSS 163
Qy      83 -----PVDVAGGLDN----- 94
Db      164 SGPVTVPASAGTTTCQPCQCQWTEFDEYDKSEQLGGDVESYDKIRAAGHLCOQPKDI 223
Qy      95 -----GDFKV----- 99
Db      224 EQAESFFNWTIAQVQKQVCHDVHGLVCRNWEQGVFKMCMYNYRIVLCCSDDHCRGA 283
Qy      100 TTP-----DSTTQHNV-----STSPKITTAKPINP----- 126

```

```

Db      284 TTPPTTELEATATTTTQALFSTFQPTSSPGLTRAPPASTTAVPTLSEGLTSPRYTSILG 343
Qy      127 -----RPSLPN----- 133
Db      344 TATTGGPQSAGSTBTPVGVATSTLPTRSALPGTGLGTRWPSQPTPLAPTTWATSRA 403
Qy      134 -----SDTSKE---TSL--TVNKETTVETKTTTTNKQTSPTDGKEKTSKAKET----- 176
Db      404 RPTGASTASKEPLTSLAPLTLSELSTQASTSPRTETTMSPLTNTTTSQGTTRCQPK 463
Qy      177 -----QSIE----- 180
Db      464 CEWTEWFOVDFTPSGVASGDMETFENIRAAAGKCMWAPKSIIECRANPEVEIDQGVQL 523
Qy      181 -----KTSAKDLAP-TSKYL 194
Db      524 TCSLETGLTCRNEQDGRFNMCMYNNRVLCDDYSHCEPSTLSTSTATSPSTPGTTWIL 583
Qy      195 AKP-----TPKASTTKGPAALTTPKETTP----- 218
Db      584 TKPTTTATTTASTGSTATASSTQATGTHVSTTATTPVTSSK-ATPFSSEGTATALPA 642
Qy      219 -----TPKEPASTTPKEPTTIK----- 238
Db      643 LRSTATPTATSTAIPTSSSLGTTWRLSQTTTPMATMTATPSSTPETHVHSTVLTYTA 702
Qy      239 -----SAPTPP----- 244
Db      703 TTTGATGVATPSSTPGTAHTTKVLTITTTGTATPSSSPGRARTLPVWLITTTTPTTRG 762
Qy      245 -----KEPAPTTT----- 252
Db      763 STVTPSSIPGTHPTVLTITTTTATGSMATPSSSTQTSIGTPPSLTATTTATGSGTT 822
Qy      253 ---KSAP--TPKPEPAPTTT-KEPAPT---TPKEPAPTTPKEPAPTTKS-----APTTP 298
Db      823 NPSTPGTTPIPVLTATTTAATSTVTPSSALGTHTPPVNTTATTHGSLSPSSP 882
Qy      299 -----KEPAPTTPKAPPTT---PKEPAPTTPKEPTTTPKAPA 334
Db      883 HTVCTAWTSATSGILGTHITEPSTGTSHTPAATGTTQHSPTALSSPSPSSKTTESPPS 942
Qy      335 PTTKEPAPT---TPKEPAPTAPKAPPTT---PKBP--APT----- 368
Db      943 PGITPGHTTATSKTTATATATPSKTRTSTLLPSQTSAPITVVTVMGCEPOCAWSEWLDYS 1002
Qy      369 ----- 368
Db      1003 YPMPEPSGGDFDTYSNIRAAAGVACEOPLGLECRQAQOPGVPLRELQGVVECSLDGLVC 1062
Qy      369 -----PKEPAPTTPKES----- 381
Db      1063 RNREQVKFKMCMYNYRIVFCNVGHCPSTPAUSSTATPSSTPGTWTWLTTELTTATTE 1122
Qy      382 -----PTTPKEPAP-----TTTKSAPT----- 398
Db      1123 STGSTAPTSTLRATAPPKVLTTTATTTPTVTSKATPSSSEGTATALPALRSTATPTAT 1182
Qy      399 ----- 398
Db      1183 SVTPIPSSSLGTTWRLSQTTTPTATMTATPSSTPETAHTSTVLATATTTGATGVAT 1242
Qy      399 -----TKPEPAPTTT---KSAPTTPKBP-----SPTT----- 422
Db      1243 PSSTPGTAHTTKVTTTTTGTATPSSSPGALAPPVMI STTTTPTTRGTVTPSSIPGT 1302
Qy      423 -----TKEPAPTTPKEPA 435
Db      1303 THTATVLTITTTTATGSMATPSSSTQTSPTGTPSLTTTATTTATGTTNPSSTPGTRPI 1362
Qy      436 P-----TPPKKAPV---TPKEPAPTTPKEPAPTTT----- 463
Db      1363 PPVLTITTTATTPAATSTVTPSSALGTHTPPVNTTATTHGSLSPSSPHTVRTAWTSAT 1422

```

```
QY 464 -----KKPAPT-----APKEPAPTTP----- 479
Db 1423 SGTGLGTHITPSTNGTSHTPAATGTTQHSPTALSSHPSSRTTESPPSGTTPGHHTA 1482
QY 480 --KETAPTTPKK-----LTPTTPKLAAPTTP----- 503
Db 1483 TSRTTATATPSKTRTSTLLPSPTS-APITVTWVGCEPQCAWSEWLDYSYMPGPGSGD 1541
QY 504 -----EKP----- 506
Db 1542 FDTYSNIRAGGAVCEQPLGLECRAQAGVPLRELQGVVECSLDGLVCRNREQVGKFK 1601
QY 507 -----APTPEELAPTTPBE----- 524
Db 1602 MCFEYRVEFCNYGHCPSPATSSPTATPSPTGTTWILTEQITAAATTTATGTAIPSS 1661
QY 525 TTPEEPAP-----TTP-----KAAAPNTPK-----EPAPTTP-- 550
Db 1662 TPGTAPPKKVLTQATTFATSSKATSSSPRTATTLPVLTSTATKSTATSFPIPSSTL 1721
QY 551 -----PKEP-----APTTPKEPAPTTPKETAPTTP-- 574
Db 1722 GTTGTSONRPHPMATMSTIHPSTPETHSTVLTAKATTATATSSMSTPSPGTTWI 1781
QY 575 -----PKGAPTTP-----LKEPAPT----- 589
Db 1782 LTELTTAATTAAALPHGTSPSTPGTTWILTEPSTATVPTGSTATASSTRATAGLKV 1841
QY 590 -----TPKK-----PAKELAPT----- 602
Db 1842 LTSTATTPVVISSRATPSSPGTATALPALRSTATTPTATSVTAIPSSSLGATWRLSQT 1901
QY 603 -----TTKETSTST----- 611
Db 1902 TTPATMSTATPSSTPETHSTVLTATTTRTGVATPSSTPGTAHTTKVPTTTTGTG 1961
QY 612 -----SDKPAPTTP----- 619
Db 1962 TATPSSSPGTALTPPVWISTTTTPTTRGSIPTBSSIPGTHATVLTITTTTIVATGSMAT 2021
QY 620 -----PKGAPTTPKEPAPTTP----- 635
Db 2022 PSSSTQSGTPPSLTTATTATITATGSTINPSSPGTTPPPVLTATTTPAATSSIVTSP 2081
QY 636 -----PKEP-APT-----TPK 645
Db 2082 SALGTHTPPVNTTATTHGRSLPPSSPHVTPTAWTSAATSGILGTHITPSTGTSHTPA 2141
QY 646 GTAPTTP-----LKEPAP-----TTPKKPAPKELAPTTPKGTSTTSDDKPAPTTPKET----- 692
Db 2142 ATTGTTQPSPTALSSHPSSRTTESPPSP-----GITTPGHTGRSTTATATPSKTRTST 2197
QY 693 -----APT-----TTPKEP----- 700
Db 2198 LLPSSFTSAPTITVTVTGCEPQCAWSEWLDYSYMPGPGSGDFTYSNIRAGGAVCEQP 2257
QY 701 -----AP----- 702
Db 2258 LGLECRAQAGVPLRELQGVVECSLDGLVCRNREQVGKFKVCFNYEIRVFCNNGHCP 2317
QY 703 TTPKKPAPTTPPTTPPTTP----- 720
Db 2318 STPATSSATPSSTPGTTWILKLITTTATTTSTGSTATPSSTQGPAGTPHVTATTATP 2377
QY 721 -----SEVSTPTTP----- 729
Db 2378 TVTSSKATPPSPGTATALPALRSTATTPTATSFATAIPSSSLGTTWRLSQTITPMATMS 2437
QY 730 -----KEPTTIHKS----- 738
Db 2438 TATPSSSTPETHSTVLTATTATGATSVATPSSTPGTAHTTKVPTTTTGTFTVTPSSS 2497
```

```
QY 739 -----PDEST-- 743
Db 2498 PGTARTPPWISTTTTTPTTSGSTVTPSSIPGTHHTPTVLTITTPQVATGSMATPSSSTQT 2557
QY 744 -----PELSAEPPTPKALEN----- 757
Db 2558 SCTPPSLITTATTITANGSTTNPSSTPGTTPIPPILTITATTTPAATSTVTPSSALGTH 2617
QY 758 -----SPKEP-----GVPT 766
Db 2618 TTPVNTTATTHGRSLSPSSPHVTARTATWSATSGTLGTHITPSTGTSHTPAATTGTTT 2677
QY 767 TKTPAAATKPEMT-----TTAKDKTTE----- 787
Db 2678 TSTPALSSPHSPSSRTTESPPSPGTTTTPGHTTATSRITATATATPSKTRTSTLLPSQTSAPI 2737
QY 788 ----- 787
Db 2738 TTVVTTGCEPQCAWSEWLDYSYMPGPGSGDFTYSNIRAGGAVCEQPJGLECRATAQP 2797
QY 788 ----- 787
Db 2798 GVPLGELQGVVECSLDGLVCRNREQVGKFKMCFNYEIRVFCNNGHCPSTPATSTAMP 2857
QY 788 -----RDLRTTPETT-----TAAPKMTVKETA-----TTTEKTIESKI-- 819
Db 2858 SSTPGTTWILTELTTTATTASTGSTATPSSTPGTAPPKVLTSPTATTPTATSSKATSSS 2917
QY 820 ---TATTQV-TSTTTQDT---TPFKITLTKTLLAPKVTTTK-KTITTEIMNKPEE- 869
Db 2918 SPRTATLPLVLTSTATKSTATSVTPIPSSSLGTLGTLPSQTTTPPVATMSTIHPSSPETT 2977
QY 870 -----TAKPKORATMSKAT-----TPKPQK----- 889
Db 2978 HTSTVLTTKATRATSTSTPSSTPGTTWILTELTTAATTAGTGTATPSSPTGTTWIL 3037
QY 890 ----- 889
Db 3038 TELTTTATTASTGSTATSLSTPGTTWILTEPSTTATVAPPSTATASSTQATAGTPHV 3097
QY 890 -----PTKAPKKTSTKKPKT---MPVRVKPKTTPTPRKMTSTMPELNPTS-----R 933
Db 3098 SITATPTVTSSKATPSSSPGTATAPLRSTATTPTATSFATAI-----PSSSLGTTWR 3152
QY 934 IABAMLQTTTRNQTPNSKLVEVNPKSADAG-----GAEGTTPHMLLRPHVEMPE 983
Db 3153 LSQTTTPTATMTATPSS---TPETVHTSTVLTITATTGATGS-----VATPS 3198
QY 984 VTPDMDYLPRVP-----NOGIINP 1003
Db 3199 STPGTAHTTKVPTTTTGTATPSSSPGTALTP 3231
```

## RESULT 8

S48478  
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (*Saccharomyces cerevisiae*)  
N:Alternate names: extracellular glucoamylase; mucin-like protein MUC1; protein YIR019c  
C:Species: *Saccharomyces cerevisiae*  
C>Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text\_change 09-Jul-2004  
C:Accession: S48478; A26877; B26877; S27281; JG6123  
R:Rowley, K.  
submitted to the EMBL Data Library, October 1994  
A:Reference number: S48478  
A:Accession: S48478  
A:Molecule type: DNA  
A:Residues: 1-1367 <ROW>  
A:Cross-references: UNIPROT:P08640; GB:Z47047; EMBL:Z38061; NID:G603997; PID:G763364; GS  
J.Yamashita, I.; Nakamura, M.; Fukui, S.  
J. Bacteriol. 169, 2142-2149, 1987  
A:Title: Gene fusion is a possible mechanism underlying the evolution of STAL.  
A:Reference number: A91831; MUID:87194600; PMID:3106330  
A:Accession: A26877  
A:Molecule type: DNA







```

QY 111 -----KVSTPKITTAKEINPR-----PSLPPNS- 134
Db 466 ISSHGAFTTTGEQIIEGSGIQTSSGSGSIITVRGLIPKQVELITISGRFACTSCYPNSE 525
QY 135 -----134
Db 526 CVGGYCEVSGYGNALVGCBDIDECITEICNIEANECVNLIGFVCCNPTNATHDDCI 585
QY 135 -----DTSKETSLTV---NKETTIVETKETTINKOTS----- 163
Db 586 DFLTKVIYAMIIFLLKGLIEITKBLHWIGNEEDIV---VATRSNHSTSOQLITQV 642
QY 164 -----TDGK-----EKTSAKETQSIIEKTSADKLAPTSKVLAKPT---P 199
Db 643 VQOSRNFSTGQIILTRGVSSGEAVTQTTDADEFL--EISAADLAGSGSITLPTLLEP 700
QY 200 KAE-----TTTKGPALT-TPK----- 214
Db 701 KIEGSGKASGVVWTEDEGEDEDLMBEGSGSWTTINGTIGTSPRSEGTIRVRITLIG 760
QY 215 ---EP-TPTTP-----KE----- 223
Db 761 EDGEPEATATPGISAPDKTGEKSTESDGEKLTVEKDGEKAEQSSGSATSSGKKSEATS 820
QY 224 ----- 223
Db 821 GSSSSSAKSGTGFASGSGASSSSGSGVSGESSVSTESGSGFTSSSGSVGSEATGST 880
QY 224 -----PASTTPKEPTTT-----IKSAPTTPEKAPATTTK----- 253
Db 881 GVDGSSGKPSKSTSEKLPFTKNGEKSPISGSDTTGKESSEETTSRKPIEGSDSLTEGS 940
QY 254 -----SAPTTPEK-APTTP----- 267
Db 941 GSGEWFTGSKGHFESGSKVSVTSGKPTQSGAEGSGGPKVKGPGAPBITTDGBESS 1000
QY 268 -----KEPAPTTPEKAPATTT----- 283
Db 1001 TSTGDKSGGKPADKSDNKNVPGTKDNPDITTDGEDSTSETSGEGQPKGSKQPPGDK 1060
QY 284 ---KEPA-----PT-----TTKSA----- 294
Db 1061 GSEVKKTEPVDGPNLSGTKGKSNVPLKPTDLPERGSGILTTSGGKNSFTBEHGTCLR 1120
QY 295 -PTTPKEPAPTTP-----KKPAP-----TTP----- 314
Db 1121 LPKTEKSSETPOLGLBISAGKKPEPDGTSKEVGLLEILWESTTPGSTLDSVGLLEI 1180
QY 315 -----KEP-----APT 320
Db 1181 SGSDLTAKKPHVEIEGSGTGDEEITATTRDVSKSTKKPRVEVDGDNGETSGVDGKPT 1240
QY 321 TPKEPPTTPKEPA---PTTPKEPAPT-----TPKEP----- 348
Db 1241 TP-APTSSASBSTSRIPITTEASPEGSGGAGVPESPDPGSGESSTAPDGVSPTSST 1299
QY 349 APTAPKAPAPTTP---KEPAPTTPK---EPAPTTPKEPPTTPKEPA---PT--- 391
Db 1300 APEVPTTSASTPDAVEESGIFSTSKPTAEPLLETTAPSTEVTSPEGSGTEESTLPPEGS 1359
QY 392 ---TKSAPTTTKEPA-----PTTKSAP----- 412
Db 1360 GESITSSAPTIV--EPATVLPQNRNEKPEPTKDTFALPTTTTCAPQANDSSVENTKCTSS 1417
QY 413 -----TTPKEP----- 418
Db 1418 DECGLDALCERRTGVCRCEPFGEPAGPKKSCVDVBECATGDHNCHESARCQNVGVYACF 1477
QY 419 SPT-----TTK 424
Db 1478 CPTGFRKADGSCQDIDECTEHNSTCCGANAKCVNKPVTGYSCECENGFLGDGYQCVPTTK 1537

```

```

QY 425 EPAPTTPK-----EPAP----- 436
Db 1538 KPCDSTOSSKSHCHSESNNSCVDTVDSVECKECMGGYKKGKVCEDINECVAEKAPCSL 1597
QY 437 ----- 436
Db 1598 NANCVMNNGTFCSCQKQYRGDGFMCCTDINECDERHPCHPAECTNLEGSFKCECHSGFE 1657
QY 437 -----TTP----- 439
Db 1658 GDGIKKCTNPLERSCEDVEKFCGRVDHVSLSVRIYNGSLSSVCECPGPFPEKESNCV 1717
QY 440 ----- 439
Db 1718 DIDECEBSRNNCDPASVCVNVTEGSRCECAEGYEGGVCTDIDECDRGMAGCDSNMAC 1777
QY 440 ----- 439
Db 1778 INRMGSCGCKMAGYTGDGATCIKIEEPKSDKTACTDEWSRLCELEKKQCTVDEBEVPQ 1837
QY 440 ----- 439
Db 1838 CGACLPGHPIGTQOSLQISGLCAQKNDCKNHAECIDIHPDSHFCSGPDGFIGDGMICD 1897
QY 440 ----- 439
Db 1898 DVDECNNAGCDDENTKCENTIGSPNCVCLGFKVKVDEKCVDEKKQPNREKIEIDENS 1957
QY 440 ---KKP-----APTTPKEPAPT-----PK 456
Db 1958 SSSNSGOEKTTKGIYSSTSATSESTAEPHVTTSISSTSTKDMTSSKSPENVMTSSE 2017
QY 457 EPAPTTPKAPAPKEPAPTTPKET---APTTPKKLT-----PTTPE--- 496
Db 2018 SPEVSTSSSKSTTASETTVSSSTPSSSSSEAPLTSPPATTTVEVITESVSKSTTPKESSS 2077
QY 497 ---KLAPTPR-----KPAPTTPELAPTTPEEPTPTTPPEPAPTTPKAAAPNTPKBP 546
Db 2078 EITVLUSSKSPVETESSVSSPSTPS---TTSQSVTSTVPETSKSTLSSEAPVITSTSP 2133
QY 547 ---APTTPKEPAPTTPKEPAPTTPKET---APTTPKGT---APTTLKEPAPT 590
Db 2134 TEVHTSSEIKPSLASISITGDTNSTITPSSLASVKSSTAPEGTSASVAPVLUSSUSPDV 2193
QY 591 PKKAPKELAPT-----TTKEP-----TSTTSKAPAPTTPKGTAPT 627
Db 2194 -SQPSTKTFTDATESSTVQASSETSSGTSVKSTSEPESHVTKLSITSSNPSSVPVTSFKST 2252
QY 628 PKEPAPTTPKEPAPTTPKG-----TAPTTLKE 654
Db 2253 PTVPEST--EQPTSTTPSGQSLTPMNSNSEVLTTSPEHVLSSSLSPDVSQSSTTPNNLSE 2310
QY 655 PAPT-TPKKPAPKEL---APTTPKGT-----STT-----S 681
Db 2311 SSTVETPKTSSEVLSNSEPSTTEAPTLLSPDILSITTNLQSSTSTVEDSEISSENS 2370
QY 682 DKP-----APTTP-----KETAPTTP----- 697
Db 2371 EKPTSAPELVTSVTHVASSSPDVPTESSEPDDLGSSTENIPEASSKQTSISTPTPDDT 2430
QY 698 ---KEPAPTTPKPP-----APTPE----- 714
Db 2431 TASEEPTKSTGMSPLSTTSNVLSSESTTPESSKSPVSSSTEGISVVVSTSTEFKVPST 2490
QY 715 -----TPPPTTSEVSTPTTTKEPT---IHK--SPDESTEPELSAEP 751
Db 2491 ISSVLEEDLTKTTPSPILIEETITASETSEPLETDSLTVSRLHELTTSSENVKSEST 2550
QY 752 PKALENSP-KEPG-----VPTT-----KTPAATKPEM--- 777
Db 2551 TSSESSKPSQBPAGILTSTVVVPTSSVSLITASEIEALISNTPTFKQGTPTTTSFKSLVK 2610
QY 778 ---TTTAKOKTHERDLRTPPTTTAAPKMTKETATT-----TEKTTESKIT 820

```

Db	2611	STTSSTVTSSEPSSTKKTIVSTTVSTTTTPEETTSLSLTLTAAPSKPTSTSTESSA	2670
Qy	821	ATTIQVTS-----TTQDTTPPKITLTKT-----TLAPKVTTTKK	856
Db	2671	PTTPAKTSTKPSNVSSSTRKSPENVETSTSQSGLSSTMSSTSEPETNAPAVTSSE	2730
Qy	857	TTTTTT-----EI	863
Db	2731	ASSTTLEENSSSTSPSSBASVKLSLFPESITSEAVTVSSRAPABITMSSESHREISTV	2790
Qy	864	MNKPBR-----TAKPKD-----	875
Db	2791	SEEPSEPEIPLSTTVSPNVVTVASSIPSEBILSVTSSTSPRVRLITGTDPDDLIVSVTV	2850
Qy	876	-----RATNSKA-----TTPKPQKPTKAPKXPTSTKKPKTPRVKXPK	913
Db	2851	SHGNRRQNTAVASSVPNSNSTPIILPSESLTTPQPPPTTTTAKPAUTSGKRGPPSIQPPA	2910
Qy	914	-----TTPTTP-----RKMTSMPELNPSTRIAEAMLOT-----	941
Db	2911	EMFTTAPPPPSNGYGEETNQBEQVIST-----TTTAPSLCSTVTCHSLATCEQST	2964
Qy	942	-----TTRNQTPNSKLVNPKSBDAGGAGET-----P	971
Db	2965	GVICIRGDFIGDGTACSKKSTADICISLPLCADKAKCDNSTRSCEDAGYIGDGYVCS	3024
Qy	972	H---MLLRPHVFWPE	983
Db	3025	HPQDCVLRDLNCSPE	3039
RESULT 12			
T25697			
hypothetical protein F16F9.2 - Caenorhabditis elegans			
C;Species: Caenorhabditis elegans			
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004			
C;Accession: T25697			
R;Fulton, B.			
submitted to the EMBL Data Library, August 1996			
A;Description: The sequence of C. elegans cosmid F16F9.			
A;Reference number: Z20071			
A;Accession: T25697			
A;Status: preliminary; translated from GB/EMBL/DBJ			
A;Molecule type: DNA			
A;Residues: 1-1229 <FUL>			
A;Cross-references: UNIPROT:Q94185; EMBL:U67956; PIDN:AAB07691.1; GSPDB:GN00028; CESP:F16F9			
A;Experimental source: strain Bristol N2; clone F16F9			
C;Genetics:			
A;Gene: CESP:F16F9.2			
A;Map position: X			
A;Introns: 35/1; 361/1; 384/1; 482/2; 517/1; 971/1; 1021/2; 1179/3			
Query Match			
Best Local Similarity 15.7%; Score 844.3; DB 2; Length 1229;			
Matches 334; Conservative 95; Mismatches 360; Indels 497; Gaps 63;			
Qy	2	AWKTLPIYLLLLSVFVIQVSSQELSCG-RCFESFERGRCDCDAOCKYKCCPDYE	60
Db	4	AW-VVSFAFLILGNVQSSLLSKTINSIGSRDPKIVKHKKNTCTCSC---KCPD--	56
Qy	61	SFCAEVKDNKQRTKKTPTKPPVVDGAGSLNDGDFKVTPTDSTT---QHNKVSTSP	116
Db	57	-----APGNPF-----DVSTTSSINNNNDVIGP	80
Qy	117	KITTAKPINRPSLPNDSKTSKTSLVNKKETT-----VETKETTNNKQTS--	163
Db	81	S-----GDSNFTGSSWFQIEATVGGQTVKSEHNIDSSVEKKVTTSTDASTNA	131
Qy	164	-TDGKEKTTSA-----KETQSIEKTS-----AKDLAPTSKVLAKP	197
Db	132	PTTGKDSITTEIITGIWVINSKSESVTMSITRFTSTLSPTELLTSPETLVSTDSSTST	191

Qy	198	-----TPKAETTTKGPALTTPKEPTPTT---PKEPASTTTKPEPTPTTIKSAPTT	243
Db	192	EQTSNDNTTEIASPMETNTTTEATTTTSVEPSVSTLASEDETTVTAIAESTTTVIAEVSTT	251
Qy	244	KPEP---APTTSKAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE	300
Db	252	TBEPTTTABSTTKKST---KAPATTEPTPTTTEE---VTTTEASTTSTTSETSETEK-	304
Qy	301	PAPTTP---KKPAPTTPKEPAPTTPKEP-TPTTPKEPAPTTPKEPAPTTP---KEPAPT	351
Db	305	---PTTPLIUNKIAGPAIGK---PETTHFFVTGTP-----NPDTEATEPFFVAKSEDKMT	353
Qy	352	APKAPAPTTPKEPAPTTPKEPAPTTPKE-----PSPTTPKEPAPTTPKKSAP-----TT	399
Db	354	LSKTAATEITQOTTEVT-DGPEKETTKNWSIEPIITTVPLIVETTTSTTASKESDGHPTT	412
Qy	400	TKEPAPT---TTKSAPTTPKEPSPTTKKE---PAPT---TKP-----EPAPTTP	439
Db	413	LKLKVTTADSDSTESATTVKPFNEETTTKSHVVPKTKGTVKVKTPKLELSFDEPTEIT-	471
Qy	440	KKPAP-----	444
Db	472	KAPHPKLLKKTYYHFVLSDNFARYSEAKENDDYNHLDYNHYREAKEPTTTTEESSTEEV	531
Qy	445	TTPEPA-----PTTPKEPAPTTPKXPAPTAPKEPA---PTTPKETAPTTPKKUTPTTPEK	497
Db	532	TTTEEPANTGNPPTTEN---PTTTEQPTSTAESTTTTALPFTTETV-----TTEE-	578
Qy	498	LAPTTPKEPAPTTPPELAPTTPPEPTTPPEEPAPTTPKAAAPNTPEPAPTTPKEPAPT	557
Db	579	--PTTAKSTATQ---KPTTQESVST---EKTSITTKA---STTEE---PTTDEPTTT	624
Qy	558	TPKEPAPTTPKETAP---TTPKGTAPTTLKEPAPTTPKDPAPKELAPTTPKETSTSTSDK	614
Db	625	T---ESTTKGATTPBELSTTSSEITTELK---ITTE-----GSTTTEEPITTAIPA	670
Qy	615	PAP-----TTPKGAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTP-----	660
Db	671	EASTGIITDEETSTSTTPTSTSTKE--IVTESAITQTSVSVESSTPQLPWRKAI	728
Qy	661	-----KKPAPKELAPTTPKGP-----TSTTSDKAPAPTTPKE	691
Db	729	VNKFKNHLYLKEKKRLLEKESSTTSGSDSSETTTVAENIDEVTTTEKEKVQVTPIT	788
Qy	692	TAPTTPKEPAPTTPKXPAPTTP---TDPPTTSE-VSTPTTTPKEPTTIHKSPDESTPELSA	748
Db	789	TEKSTTQEEETTTTTEKTTSTKTTTEKPTTSESATTTETISEPST-----TEST---	838
Qy	749	EPTPKALENSPKBPGVPTTKTPAAKPEMTTAKDKTTERDLRTPPETTTA-----	799
Db	839	-----TVDTSATTEESSTAETTTTSAB---TSETTTSESAAFIGE	878
Qy	800	-----APKMTKETATTTTEKTTESKITATTQ-VTS	828
Db	879	SPENTALQSSSQKSEENESSAABKPGARRDFVPKKHKTTPKPAETTSVAASATTTTEPIT	938
Qy	829	T---TTQDTTPPKITLTKTTLAPKVT---TTKKTITTTTEIMNKPEETAKPKDRATNSKA	882
Db	939	TEKSTTLETTPLEATTLNEVTPGPAVTCAPVDEITTLLELLSK-----INNTOI	988
Qy	883	TTPKPQ-----KPTKAPKPTSTKPKTPMRVRKPKTTTTPKMTSTM	925
Db	989	SQPKPTDISKTDALSSLLSIGLIGSFTKAPMPTI-----HTTTDAAFVTATE	1035
Qy	926	PELN-----PTSRIAEAMLQTTTTPNCTPNSKLVNPKSBDAGGAGETPHM	973
Db	1036	ASLNDGSDKKIIDBAQPTDEIRRA-----QPTN-----	1063
Qy	974	LLRHFVFMPEVTPDMVDYLPVRVNOGI	999
Db	1064	-----EMDKEMEPEKRIEORI	1080



RESULT 13

T18535  
high molecular mass nuclear antigen - chicken (fragment)  
C:Species: Gallus gallus (chicken)  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T18535  
R:Shimada, K.; Harata, M.; Mizuno, S.  
J. Cell Sci. 110, 3031-3041, 1997  
A:Title: A nuclear matrix-associated high molecular mass nuclear antigen, HMNA, of chick  
A:Reference number: Z18955; MUID:9803440; PMID:9365273  
A:Accession: T18535  
A>Status: preliminary; translated from GE/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1151 <SHI>  
A:Cross-references: UNIPROT:O57580; EMBL:D88440; NID:dl171138; PID:di025045; PIDN:BAA241

Query Match 15.3% Score 822.8; DB 2; Length 1151;  
Best Local Similarity 22.4%; Pred. No. 7.1e-10;  
Matches 280; Conservative 91; Mismatches 346; Indels 532; Gaps 54;

QY	77	KTPK-----PPVVDEAGSLDNGDFKVTTPDTSTQHNKVS-----TSPKLTT	120
Db	47	QPPKWPVIGGAPP-----PGEPTPSKPTGDADAAPKAAELTS	88
QY	121	AKPINRP-----SLPNSD-----	135
Db	89	PPFASPPDPGPKAPSGAGEAEAGTPPSQGPAGTPPPSQGAAGAPKGDGTAQPSGKSG	148
QY	136	-TSKETSITVNKETTVEK-----ETTTNKOTSTOGKEKTTSAKETOSIEK-----	181
Db	149	ADGKPAQDVPKATTAATEARPASASPTVPKATA-----EATAVTAASQSAKPAAATDAAA	204
QY	182	-TSAKDLAPTSKVLAKP-----TPKAETTTKGPALTTKPEP-----	216
Db	205	VTAASQSAKPAATVEVKPAAAVAAKEAKAVTAAAAAPKATAEAKPAPVTSPTIPCSSABAK	264
QY	217	-----TPITPKPEASTTPKEPTPTTKSAPTTPKPEBAPTTTKSAPTTPKPEAPTTTK-----	268
Db	265	PLTAASPTASKATAEAKVPATASLTMATKVTAEAKPAP-----SPSVFK-ATTDTKAVT	317
QY	269	-----EPAPTTKPEAPTTTKEBAPTTTKSAPTT-----PKPEA	302
Db	318	ATAPKAGDVKPDAVCAEAKGAPPPPOOLPKAAAAAAPTGTCLKPATAPPHGSPRANS	377
QY	303	-----PTTKKKPAPTTKPEAPTTPKBPTPTTKPEBAPTTTKEBAPTTTKEBAPTAPKX	355
Db	378	HTVTVTPNVPRAAAATVP--TAGAVPKASTGTT-----AAAQQQVP---K	420
QY	356	PAPTTKPEAPTTKPEAPTTTKESPTTPKPEBAPTTTKSAPTTTKEBAP-----TTT	408
Db	421	AAPVTPSPQOAVPR-----AATAAAAPVTFQOP---VTKAAATTNATPPQPPKAAATTT	473
QY	409	KSAPTTKPEPSPT--TTKEBAPTTKPEP-----APTTPKKPAPTTTKEBAPTT	454
Db	474	TATPVTQQOPIPKAGTDAAAPPVAVPKAPDGGRAATPGVFNAAATDPOKPPPTPOSVEGAVT	533
QY	455	PKPEAPTTTKKPA-----PTAP-----KEBAPTTTK-----ETAPTT	486
Db	534	EPKQPRAAAPPSNEATPAVPSPSNLKSPLTIPKPVFLMALTPQPVTAQMWTLQAAIK	593
QY	487	PKKLTPITPEKLAPTTTPEKAPTTPEELA-----PTTP-----EETPTTTP--	527
Db	594	PSPIVPKASPK-ALMTTPPPPPGLPRALAAAKLLGLPSPFVASAMHAKVTTPRLPASVFP	652
QY	528	--EBAPTTTPKAA-----	538
Db	653	MAASPASLGPDAARVALATNAASPGAKPEAAGNGTLMAPMGAANTQMAPIGAAGAAQTA	712
QY	539	-----	
Db	713	PMGAAATHVSPMGAGGATQMSPTGAANTHMSPIGAGGATQMSPMGAANTQMSMGATTTQ	772
QY	552	KEP---APTTPKEP-----APTTPKETAPTTPKGT---A	579



Db 972 -----DTRP-----APGSTAPPAGHGVTSAPDTRPA 996  
QY 987 FQKPTKAPKPTSTKKPTMPKVRKPTTPKMTSMBELNP----- 930  
Db 997 P--GSTAPPAGHGVTSAPDTRP-----APGSTAPPAGHGVTSAPDTRPAGLSTAPPVHNVTSAS 1051  
QY 931 --TSRIADAMLO--TTTPRNPOTPNK-----LVEVNPKSEDAGGAGET 970  
Db 1052 GSASGSASTLVHNGTSARATTPASKSTPSPISPHSHSDTPTTILASHSTKTDASSHTHTV 1111  
QY 971 PHMLLRPHVFEVPT-----PMDY 990  
Db 1112 PPLTSSNHSTSPQLSTGVFFFLSHFISLNQFNSSLEDPSDY 1154

RESULT 15  
B48666  
cell proliferation antigen Ki-67, short form - human  
C:Species: Homo sapiens (man)  
C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 15-Mar-2004  
C:Accession: B48666  
R:Schlueter, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Gerdes  
J. Cell Biol. 123, 513-522, 1993  
A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, ubiqui-  
ins.  
A:Reference number: A48666; MUID:94043435; PMID:8227122  
A:Accession: B48666  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-2897 -SCH>  
A:Cross-references: EMBL:X65551  
C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat  
F:29-91/Domain: kinase interaction domain homology <KIH>

Query Match 15.1%; Score 811.2; DB 2; Length 2897;  
Best Local Similarity 15.6%; Fred. No. 4e-09;  
Matches 383; Conservative 128; Mismatches 404; Indels 1548; Gaps 92;

QY 22 VSSQELACK-----GRCFESFERGRCDGD 46  
Db 382 ISNQKWFDEDLGSLAEMFKTPVKEQPQLTSTCHIALNSENLLGQFQGTDSGE----- 437  
QY 47 AQCKYDKCCPDYESCAEVKDNKQRTK---KPTPKPPV---VDEAGSGLDNGD--- 96  
Db 438 -----PLLPTESEFGNVFFSAQNAAKQPSDKCSASPLRRQCIRE-----NGNVAK 484  
QY 97 -----FKVT---TPDTSIQHNKYST---SPKTIITAPINRP----- 128  
Db 485 TPRNTYMTSLTKTSDTETEPSKTVSTVNRSGRSTEFNRNIQKLPVESKSEETNTEIVEC 544  
QY 129 ----- 128  
Db 545 ILKRGQKATLLQORREGMEIERPFETYKENIELKENDEKMKAMKRSRTWGQKCAPMSD 604  
QY 129 -----SLP-----PNSDTSKETSITVKN-----ETIVETK- 153  
Db 605 LTDLSLPDTMLMOTARGONLIQODHAKAPKSEKGIKMPQCSLSQLEPINTPHITQ 664  
QY 154 -----ETTTTNKQISTDGE----- 168  
Db 665 QLKASLGKGVKBEALLAVGKFTRTSGETHTRREPAGDGKSIRTFKESPKQILDPAARVT 724  
QY 169 -----KTSASKETQSIEKTSAKDLAPTSKVLAKPTKAE----- 202  
Db 725 GMKKWPRTPKEAQSLE-----DLAGFKELFQTPGSEESMTDEKTTKIACKSPPPESVD 779  
QY 203 ---TTTK-----GPALTTPKEP-----TPT--- 219  
Db 780 TPTSTQMPKRSIRKADVEEFLALRLKLTSPSAGAKMLTPKAGDEKDIKAFWGPVQKL 839  
QY 220 -----TPKEPAST-----TPKEP----- 232  
Db 840 DLAGTLPGSKROLQTPKEKAQALDLAGFKELFQTPGHTTEELVAAGKTTKIPCDSPQSDP 899

QY 233 --TPTTIKSAFT-----TPKEPAP----- 249  
Db 900 VDTPTSTQRPKRSIRKADVEGELLACRLNMPESAGKAMHTPK---PSVGEEKDIIIFVGT 956  
QY 250 -----TTTKSAFTTPKEPAPT-----TTKEPAPT 273  
Db 957 PVQKLDLTENITGSKRRPQTPKEBAQALDITGFKELFQTPGHTTEEVAAGKTTKMPCES 1016  
QY 274 TPKEPAPTTPKEPAPTTPKSAFTTPKE----- 300  
Db 1017 SPPEADT-----PTSTRQPKTPLEKRDVQKLSALKLTQTSGETTHTKVPGGSDX 1070  
QY 301 -----PAPTTPKAPAPTTPKEPA-----PTTPK- 323  
Db 1071 SINAFRETAQOKLDPAASVTGSKRHPKT-KEKAQPLEDLAGWKELFQTPVCTDPTTHEK 1129  
QY 324 -----EPTPT-TPKEPAPTKE-----PAPT----- 343  
Db 1130 TTKIACRSQDPVDPTPTSSKPSKRSIRKVDVEEFFFALRKRTPSAGKAMHTPKPAVSGE 1189  
QY 344 -----TPKEPA----- 349  
Db 1190 KNIYAFMGTPVQKLDLTENITGSKRRLOTPKEKAQALDLAGFKELFQTRGHTTEESMTND 1249  
QY 350 -----PTAPKKPA-----PTTPKEPAPT- 367  
Db 1250 KTAACKSSQPDLDKNPASKRRRLKTSGLGKGVKEALLAVGKLTQTSGETTHHTHTPTG 1309  
QY 368 -----TPKEPA-----PTTKKPSFP 382  
Db 1310 DGSKMAFESPQILDSASLTGSKRQLRTPKGSVPEDLAGFIELFQTPSHTK- SM 1368  
QY 383 TTPK-----EP-----APTTPKSAP----- 397  
Db 1369 TNEKTTIKSVYRASQPDVLTPTSSKQPKRSIRKADTEEFALRQTPSAGKAMHTPKP 1428  
QY 398 -----TTTKEPA-----P 405  
Db 1429 AVGEKDIINTFLGTPVQKLDQPGNLPGSNRLQTRKEKAQALBELTGFRELFPQTPCTDNP 1488  
QY 406 T---TTK---SAPTTKPEPSPTTKE-----PA----- 427  
Db 1489 TADEKTTIKLCKSPQSDPADPTNTKQPKRSIRKADVEEFLAFKRLTPSAGKAMHTP 1548  
QY 428 -----PTTPKEPAPT-----TP--- 439  
Db 1549 KAAVGEKDINTFVGTPVEKLDLGNLPGSKRRPQTPKEKAQALDLAGFKELFQTPGHT 1608  
QY 440 -----KPAFTTPKEPA----- 451  
Db 1609 EESMTDDKITEVSCSPQDPVKPTTSSKQRLKISLGKGVKEEVLVPGKLTQTSKITTQ 1668  
QY 452 -----PTTPKEPA----- 459  
Db 1669 THRETAGDGSIKAFKESAKOMLDPAVYGTGMRWRPRTPKERQAQSLDLAGFKELFQTPD 1728  
QY 460 -----PTTKKPAFTAPKEPAP---TTPKET---APTTP-----KKLTPTT--- 494  
Db 1729 HTEBSTDDKTTKIACKSPPPESMDPTSTRRRPKTPLGKRDIIVEELSALKQLTQTHTD 1788  
QY 495 -----BEKLAP---TTPKEPAPTTP-----BELA----- 515  
Db 1789 KVPGEDKGINVRETAQOKLDPAASVTGSKRQPTPKGAQPLEDLAGKELFQTPVCT 1848  
QY 516 ---PTTPESPTTPPEPAPT----- 533  
Db 1849 DKPTTHEKTTIACKSPQDPVGTPTIFKPSKRSIRKADVEEESLALRKRTPSVGKAMD 1908  
QY 534 TPKAAA-----PNTPKPEPA----- 547  
Db 1909 TPKPAGDEKDKAFMGTPVQKLDLPGNLPGSKRWQTPPKAQALEDLAGFKELFQTPG 1968

QY	548	---	PTTPKBPAP	TTPKET	---	AP	572
Db	1969	TDRTD	TKTACKSPQ	DPDVT	PASTKQRPKRNLRKADVEE	BFLALRK	TPSAGAM
QY	573	TTPK	---	---	GTAPT	TLKEPAPT	PKKPAP
Db	2029	DTTKPAVSD	EKNINTV	ETVQKLD	LGNLPGSKRQ	---	POTPKKAEALD
QY	602	TTTKEPT	TTSDK	---	PAP	---	---
Db	2087	TPGHT	EESMTDK	ITEVCSK	PQESFKTSR	SSKQRLKIPLVKVDMKEE	PLAVSKL
QY	618	---	---	---	---	---	---
Db	2147	GETTQ	THEPTGDS	KIKAFKES	PKQILOPAA	SVTGRSQRLRTRK	KARALD
QY	641	---	---	---	---	---	---
Db	2207	FSAPG	THEESMT	IDKNTK	IPCKSPPEL	TDATATKRCP	---
QY	673	TKGPT	TTSDKPA	---	---	---	---
Db	2266	TSQSTH	THEKPA	SGDEG	IKVLQKRAKK	PNVEEPPSR	RRPRAPKEKQ
QY	692	---	---	---	---	---	---
Db	2326	LSETSGH	QBSL	TAGATK	IPCSPPLEV	VDVTASTK	EHKLRTVQKVQKEE
QY	740	DESP	ELSAEPTP	---	---	---	---
Db	2385	QTSGETT	DADKEP	AGEDG	IKALKESAK	OTPAAPAA	SVTGRSR
QY	768	KTRAA	---	---	---	---	---
Db	2445	KOPAGH	EEESMT	---	---	---	---
QY	819	ITATTQ	VTSTTQ	---	---	---	---
Db	2502	LTQTS	GETTHTD	KEVGE	GKTKAFK	---	---
QY	873	PKDRA	---	---	---	---	---
Db	2549	PKKRAQ	PLEDLAS	FQELSQT	PGHTEEL	ANGAADSFT	SAPKQTPDSG
QY	896	KPT	---	---	---	---	---
Db	2609	VEPVGD	VSTRDP	VKYSQ	SKNSNTSL	PLPFKRG	GKGDSVTG
QY	900	TKPKT	MPVRVR	KPTTPT	---	---	---
Db	2669	SKQKVA	PRARKS	ESPEV	VIWMKRS	LRTS	AKRIEPAE
QY	953	LVEVNP	KSEDA	GAGB	ETPHMLL	---	---
Db	2729	GISLSR	RQDKTEA	EQITE	QVFLA	ERIEIN	NEKKP
QY	1001	INP	1003	---	---	---	---
Db	2775	QNP	2777	---	---	---	---

Search completed: October 13, 2004, 11:59:33  
Job time : 58.4912 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 109.889 Seconds  
(without alignments)  
4449.477 Million cell updates/sec

Title: SEQ1-F  
Perfect score: 7276  
Sequence: 1 MAWKTLFIYLLLSVFIQ.....ARAITRRSGQTLSKWNVNC 1363

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.1

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq\_23Sep04:.\*  
1: geneseqp1980s:.\*  
2: geneseqp1990s:.\*  
3: geneseqp2000s:.\*  
4: geneseqp2001s:.\*  
5: geneseqp2002s:.\*  
6: geneseqp2003as:.\*  
7: geneseqp2003bs:.\*  
8: geneseqp2004s:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7261.9	99.8	1404	2	AAR26049
2	7261.9	99.8	1404	4	AAB29773
3	7261.9	99.8	1404	4	AAB60568
4	7261.9	99.8	1404	8	ADM98014
5	7258.9	99.8	1415	4	AAB32262
6	7232.9	99.4	1404	7	ADK65839
7	7043.7	96.8	1320	7	ADK65819
8	6788.7	93.3	1270	8	ADK67911
9	6774.6	91.1	1311	8	ADK67912
10	6685.9	91.9	1299	4	AAM24322
11	3552	48.8	902	4	AAB29778
12	2929	40.3	551	4	ABU53253
13	2920	40.1	546	4	ABU53252
14	2850.9	39.2	538	5	AAO18834
15	2757	37.9	513	4	ABU53254
16	1954.5	26.9	452	2	AAB80041
17	1802.7	24.8	472	4	AAB60569
18	1401.6	19.3	292	5	AAU11261
19	1371	18.8	5179	4	AAM24516
20	1371	18.8	5179	6	ABP55365
21	1371	18.8	5179	6	ABO07258
22	1371	18.8	5179	7	ADP48091
23	1371	18.8	5179	7	ADP44998
24	1196.3	16.4	8991	6	ABU08487
25	1190.4	16.4	1664	2	AAW43106

26	1109.2	15.2	5703	8	ADL23265	Adl23265 Human MUC
27	1106.6	15.2	1795	4	ABB69806	Abb69806 Drosophil
28	1100.4	15.1	4315	5	ABP43908	Abp43908 MUC5B par
29	1099.6	15.1	717	4	ABU53144	Abu53144 Human tes
30	1097	15.1	214	4	ABU53255	Abu53255 Human tes
31	1077.4	14.8	763	3	AAG38942	Aag38942 Arabidops
32	1069.7	14.7	2284	4	ABB71434	Abb71434 Drosophil
33	1044.2	14.4	1049	4	ABB61364	Abb61364 Drosophil
34	1024.8	14.1	2112	4	ABB60403	Abb60403 Drosophil
35	1024.2	14.1	4498	4	ABB58595	Abb58595 Drosophil
36	999.5	13.7	2768	4	ABB68397	Abb68397 Drosophil
37	993.4	13.7	1538	7	ADI21202	Adi21202 Novel hum
38	992.9	13.6	1460	7	ADN39110	Adn39110 Cancer/an
39	982.7	13.6	770	4	ABU53141	Abu53141 Human tes
40	987.6	13.6	778	4	ABU53143	Abu53143 Human tes
41	980.2	13.5	4873	6	ABO14747	Abol14747 Novel hum
42	980	13.5	1370	7	ADI21666	Adi21666 Novel hum
43	974.4	13.4	692	4	ABU53155	Abu53155 Human tes
44	964.1	13.3	717	4	ABU53150	Abu53150 Human tes
45	964.1	13.3	717	4	ABU53149	Abu53149 Human tes

ALIGNMENTS

RESULT 1  
AAR26049  
ID AAR26049 standard; protein; 1404 AA.  
XX AC AAR26049;  
XX DT 25-MAR-2003 (revised)  
XX DT 02-FEB-1993 (first entry)  
XX MSF precursor.  
XX KW Megakaryocyte colony stimulating factor; secretion signal; meg-CSF;  
XX KW stability; proteolytic cleavage; adhesion; alternative splicing.  
XX OS Synthetic.  
XX Key Location/Qualifiers  
FT Region 1..26 /label= Exon\_I  
FT Region 26..67 /label= Exon\_II  
FT Region 67..107 /label= Exon\_III  
FT Region 107..157 /label= Exon\_IV  
FT Region 157..200 /label= Exon\_V  
FT Region 200..1141 /label= Exon\_VI  
FT Region 1166..1212 /label= Exon\_VII  
FT Region 1213..1266 /label= Exon\_VIII  
FT Region 1266..1331 /label= Exon\_IX  
FT Region 1331..1373 /label= Exon\_X  
FT Region 1373..1404 /label= Exon\_XI  
FT Region 1411..1166 /label= Exon\_XII  
FT Region /label= Exon\_VII  
XX WO9213075-A1.  
XX 06-AUG-1992.  
XX 17-JAN-1992; 92WO-US000433.

PR 18-JAN-1991; 91US-00643502.  
PR 10-SEP-1991; 91US-00757022.  
XX (GEM) GENETICS INST INC.  
XX Turner K, Clark SC, Jacobs K, Hewick RM, Gesner TG;  
PI WPI; 1992-284660/34.  
DR N-PSDB; AAQ27223.  
XX New human mega-karyocyte stimulating factors - for treating immune  
PT deficiencies, cancer, exposure to radiation or drugs, bacterial and viral  
PT infections, etc.  
XX Claim 1, 2 and 3; Fig 1; 87pp; English.  
XX The sequence given is a full length translation from the megakaryocyte  
CC stimulating factor (MSF) precursor. The sequence covered by exons II, III  
CC and IV encodes megakaryocyte stimulating factor (MSF). This sequence is  
CC modified by the addition of an N-terminal sequence encoding a secretory  
CC leader, an initiating methionine preceding exon II and a terminating  
CC codon following exon IV. The cDNA sequence given contains sequences  
CC derived from human megakaryocyte colony stimulating factor (meg-CSF).  
CC Exon I contains the initiating methionine, and encodes a classical  
CC mammalian protein secretion signal sequence. The sequence encoding the  
CC original meg-CSF includes exons II-IV and is thought to terminate in the  
CC region between amino acid residues 134 - 147. The primary transcript of  
CC this gene may be cleaved in different ways to yield a family of mRNA's  
CC each encoding a different MSF protein. Exons V and VI are thought to be  
CC related to the activity of the factor and are also implicated in the  
CC stability, folding and processing of the molecule. These exons are also  
CC thought to play a role in the observed synergy of MSF with other  
CC cytokines. Exons V - XII are believed to be implicated in the processing  
CC or folding of the appropriate structure of the resulting factor, i.e. one  
CC or more of these exons may contain sequences which direct proteolytic  
CC cleavage, adhesion, organisation of the cellular matrix or extracellular  
CC matrix processing. Both naturally occurring and non-naturally occurring  
CC MSF's may be characterised by various combinations of alternatively  
CC spliced exons from this sequence, with the exons spliced together in  
CC differing orders to form different members of the MSF family. (Updated on  
CC 25-MAR-2003 to correct PN field.)  
XX  
SQ Sequence 1404 AA;  
  
Query Match 99.8%; Score 7261.9; DB 2; Length 1404;  
Best Local Similarity 97.1%; Pred. No. 7.3e-177;  
Matches 1363; Conservative 0; Mismatches 0; Indels 41; Gaps 1;  
  
QY 1 MAWKTLPIYLILLLSVFVIQVSSQ----- 25  
DB 1 MAWKTLPIYLILLLSVFVIQVSSQDLSSACRGCGEYSRDATCNCDYNCOHYNECCPDF 60  
QY 26 -----ELSCKGRCFESPERGECDDAQCKKYDKCCPDYESFCAEVHNPTSPSSKKAP 79  
DB 61 KRVCCTAELSCKGRCFESPERGECDDAQCKKYDKCCPDYESFCAEVHNPTSPSSKKAP 120  
QY 80 PPSGASQTIKSTTRSPKPPNKKTKVIESEBITTEHSVSENQESSSSSSSSSSSTI 139  
DB 121 PPSGASQTIKSTTRSPKPPNKKTKVIESEBITTEHSVSENQESSSSSSSSSSSTI 180  
QY 140 KIKSKNSAANRELQKLVKDNKNRTKKKTPKPPVWDAGSLDNGDFKVTTPDTST 199  
DB 181 KIKSKNSAANRELQKLVKDNKNRTKKKTPKPPVWDAGSLDNGDFKVTTPDTST 240  
QY 200 TQHNKVTSPKITTAKPINRPSLPNPSDTSKETSLTVNKEVTETKTTTTNKQTSFG 259  
DB 241 TQHNKVTSPKITTAKPINRPSLPNPSDTSKETSLTVNKEVTETKTTTTNKQTSFG 300  
QY 260 KEKTTSAKETQSIEKTSKADLAPTSKVLAKEPTPKAETTTKGPALTTTPKEPTTPKBPAS 319  
DB 301 KEKTTSAKETQSIEKTSKADLAPTSKVLAKEPTPKAETTTKGPALTTTPKEPTTPKBPAS 360  
QY 320 TTPKEPTTTIKSAPTTTKEPAPTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 379

DB 361 TTPKEPTTTIKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 420  
QY 380 APTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 439  
DB 421 APTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 480  
QY 440 EPAPTAPKKAPATTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 499  
DB 481 EPAPTAPKKAPATTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 540  
QY 500 TTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 559  
DB 541 TTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 600  
QY 560 APTAPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 619  
DB 601 APTAPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 660  
QY 620 PEEPAATTTPKAAAPNTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 679  
DB 661 PEEPAATTTPKAAAPNTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 720  
QY 680 APTTPKKAPKELAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 739  
DB 721 APTTPKKAPKELAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 780  
QY 740 TAPTTLKEPAPTTTKEPAPKELAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 799  
DB 781 TAPTTLKEPAPTTTKEPAPKELAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 840  
QY 800 KPAPTTTPEPPTTSEVSTPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 859  
DB 841 KPAPTTTPEPPTTSEVSTPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 900  
QY 860 TKTPAATKPBMTTAKDKTERDLRTTPTTAAKPKMTKETAATTTTKEPAPTTTKEPAPTTTKEP 919  
DB 901 TKTPAATKPBMTTAKDKTERDLRTTPTTAAKPKMTKETAATTTTKEPAPTTTKEPAPTTTKEP 960  
QY 920 TSITTTQDTPFKITTLTKTTLAPKVTITTKTITTTTMMKPETAKPKDRAATNSKATTPK 979  
DB 961 TSITTTQDTPFKITTLTKTTLAPKVTITTKTITTTTMMKPETAKPKDRAATNSKATTPK 1020  
QY 980 PQKPTKAPKPTSTKPKTMVRVKPKTTTTPRMTSTMPELNPTSRIAEAMLQTTTRPN 1039  
DB 1021 PQKPTKAPKPTSTKPKTMVRVKPKTTTTPRMTSTMPELNPTSRIAEAMLQTTTRPN 1080  
QY 1040 QTPNSKLVEVNPKSEDAGGAEGETPHMLLRPHVPMPEVTPDMOYLPRVPMQGIINPMLS 1099  
DB 1081 QTPNSKLVEVNPKSEDAGGAEGETPHMLLRPHVPMPEVTPDMOYLPRVPMQGIINPMLS 1140  
QY 1100 DETNINCNGKPDVGLTTLNGLTVAFRGHYFWMLSPPSPSPARRITVWGIPIPSIDTVFT 1159  
DB 1141 DETNINCNGKPDVGLTTLNGLTVAFRGHYFWMLSPPSPSPARRITVWGIPIPSIDTVFT 1200  
QY 1160 RCNCEGKTFPFKSOQYWRFTNDIKDAGYKPKIFKFGGLTGQIIVAAALSTAKYKNWPSVY 1219  
DB 1201 RCNCEGKTFPFKSOQYWRFTNDIKDAGYKPKIFKFGGLTGQIIVAAALSTAKYKNWPSVY 1260  
QY 1220 PFKRGSGIQYIYKQEPVQKCPGRRPALINPVYGEVQVRRRRRERAIQPSQTHIRIQY 1279  
DB 1261 PFKRGSGIQYIYKQEPVQKCPGRRPALINPVYGEVQVRRRRRERAIQPSQTHIRIQY 1320  
QY 1280 SPALAYQDQKGLVHNEVKVSTLWREGLPNVVTSAISLPIKRPDGYVYAKSKOQYVNDV 1339  
DB 1321 SPALAYQDQKGLVHNEVKVSTLWREGLPNVVTSAISLPIKRPDGYVYAKSKOQYVNDV 1380  
QY 1340 PSRTARAITTTRSGQTLKSKVWYNCP 1363  
DB 1381 PSRTARAITTTRSGQTLKSKVWYNCP 1404



Db 1201 RNCRGKTPFFKDSQWRTNDIKDAGYKPIFKGFGGLTGQIVAAALSTAKYKWPESVY 1260  
QY 1220 FFKRGSSIQYIYKQFPVKQCPGRRPALNPVYVGEMTOVRRRFFERAIGPSQTHIRIQY 1279  
Db 1261 FFKRGSSIQYIYKQFPVKQCPGRRPALNPVYVGEMTOVRRRFFERAIGPSQTHIRIQY 1320  
QY 1280 SPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLNPTRKPDGYDYAFSKDQYNIW 1339  
Db 1321 SPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLNPTRKPDGYDYAFSKDQYNIW 1380  
QY 1340 PSTARAITTRSGTSLSKVWYNCP 1363  
Db 1381 PSTARAITTRSGTSLSKVWYNCP 1404

RESULT 3  
AAB60568  
ID AAB60568 standard; protein; 1404 AA.  
XX  
AC AAB60568;  
XX  
DT 27-APR-2001 (first entry)  
XX  
DE Human megakaryocyte stimulating factor (MSF, CACP).  
XX  
KW Human; CACP protein; camptodactyly-arthropathy-coxa vara-pericarditis;  
KW MSF; megakaryocyte stimulating factor; synovial lubricant;  
KW chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;  
KW antiarthritic.  
XX  
OS Homo sapiens.  
XX  
XX WO200107068-A1.  
XX  
XX 01-FEB-2001.  
XX  
XX 21-JUL-2000; 2000WO-US020002.  
XX  
XX 23-JUL-1999; 99US-0145328P.  
XX  
XX 19-JUL-2000; 2000US-00145328.  
XX  
XX (UYCA-) UNIV CASE WESTERN RESERVE.  
XX  
XX Warman ML;  
XX  
XX WPI; 2001-182721/18.  
XX  
XX New composition comprising the camptodactyly-arthropathy-coxa vara-  
XX pericarditis protein in combination with an anesthetic, useful for  
XX treating osteoarthritis, or as lubricants of tissue and joints.  
XX  
XX Example 1; Page; 34pp; English.  
XX  
XX The invention relates to a method of treating osteoarthritis via the  
XX administration of a composition comprising the camptodactyly-arthropathy-  
XX coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.  
XX The composition may further comprise a local anaesthetic. The composition  
XX of the invention may be administered via intra-articular or intravenous  
XX injection. The human CACP protein is identified in the invention as being  
XX megakaryocyte stimulating factor (MSF). The gene encoding CACP protein  
XX (MSF) is located on chromosome 1q25-31, and mutations in this gene are  
XX responsible for the heritable disorder camptodactyly- arthropathy-coxa  
XX vara-pericarditis, in which patients have synovial hyperplasia without  
XX evidence of inflammation. CACP protein (MSF) acts as a synovium  
XX lubricant, and can be used to lubricate tissue and joints in the  
XX treatment of osteoarthritis. The composition may be applied to reduce the  
XX symptoms of osteoarthritis (e.g., joint pain, loss of range of movement  
XX or joint damage). The present sequence represents human megakaryocyte  
XX stimulating factor (MSF, CACP protein). Note: This sequence is not given  
XX in its entirety in figure 4 of the specification, although a GenBank  
XX accession number was given. This sequence was therefore obtained from  
XX GenBank (U70316)

SQ Sequence 1404 AA;  
Query Match 99.8%; Score 7261.9; DB 4; Length 1404;  
Best Local Similarity 97.1%; Pred. No. 7.3e-177; Indels 41; Gaps 1;  
Matches 1363; Conservative 0; Mismatches 0;  
QY 1 MAWKTLPIVLLLLSVFVIQQVSSQ----- 25  
Db 1 MAWKTLPIVLLLLSVFVIQQVSSQDLSACGRCGEGYSRDATCNCNDYNCQHMYECCPDF 60  
QY 26 -----ELCKGRCFCFESFERGECDCDAOCKKYDKCCPDYESFCFCAVHNPTSPSSKKAP 79  
Db 61 KXVCTAELSCGRCFCFESFERGECDCDAOCKKYDKCCPDYESFCFCAVHNPTSPSSKKAP 120  
QY 80 PPSGASQTIKSTTKRSPKPNKKTKVIESEBITEEHSVSENOBSSSSSSSSSSSTIW 139  
Db 121 PPSGASQTIKSTTKRSPKPNKKTKVIESEBITEEHSVSENOBSSSSSSSSSSSTIW 180  
QY 140 KIKSSKNSAANRELQKLVKONKNRTKKPTPKPPVVDVDEAGSGLDNGDFKVTTPDTST 199  
Db 181 KIKSSKNSAANRELQKLVKONKNRTKKPTPKPPVVDVDEAGSGLDNGDFKVTTPDTST 240  
QY 200 TOHNKVSTSPKIITAKPINRPSLPNSDTSKETSILTANKETTVETKETTNNKQSTDG 259  
Db 241 TOHNKVSTSPKIITAKPINRPSLPNSDTSKETSILTANKETTVETKETTNNKQSTDG 300  
QY 260 KEKTTSAKETOSIEKTSKADLAPTSKVLAKPTKAEITTKGPAITTPKEPTTPPKBPAS 319  
Db 301 KEKTTSAKETOSIEKTSKADLAPTSKVLAKPTKAEITTKGPAITTPKEPTTPPKBPAS 360  
QY 320 TTPKEPTPTTIKGAITTPKEPAITTTKSAPITTKGPAITTKGPAITTKGPAITTKGPA 379  
Db 361 TTPKEPTPTTIKGAITTPKEPAITTTKSAPITTKGPAITTKGPAITTKGPAITTKGPA 420  
QY 380 APTTTKSAPITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPA 439  
Db 421 APTTTKSAPITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPA 480  
QY 440 EPAPTAPKPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPA 499  
Db 481 EPAPTAPKPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPA 540  
QY 500 TTKSAPITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTK 559  
Db 541 TTKSAPITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTK 600  
QY 560 APTAPKPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAIT 619  
Db 601 APTAPKPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAIT 660  
QY 620 PEEPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKG 679  
Db 661 PEEPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKG 720  
QY 680 APTTPKPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAIT 739  
Db 721 APTTPKPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAIT 780  
QY 740 TAPTTLKPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAIT 799  
Db 781 TAPTTLKPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAIT 840  
QY 800 KPAPTTPETPPTTSEVSTPTTTKEPTTIHKSPDESTPELSAETTPKALENSKPEGVPT 859  
Db 841 KPAPTTPETPPTTSEVSTPTTTKEPTTIHKSPDESTPELSAETTPKALENSKPEGVPT 900  
QY 860 TKTPTAATKPEMTTAAKOKTTERDLRTTPETTTAAKPKMTKETATTEKTESKIATTTQV 919  
Db 901 TKTPTAATKPEMTTAAKOKTTERDLRTTPETTTAAKPKMTKETATTEKTESKIATTTQV 960  
QY 920 TSTTTQDPTTPKITTTLKTTTLAPKVTTTKTITTEINNKPEETAKPKDRATNSKATTPK 979  
Db 961 TSTTTQDPTTPKITTTLKTTTLAPKVTTTKTITTEINNKPEETAKPKDRATNSKATTPK 1020



QY 980 POKETKAPKPTSTKKPTMPVRVRKPTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPN 1039  
 DB 1021 POKETKAPKPTSTKKPTMPVRVRKPTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPN 1080  
 QY 1040 QTPNSKLVENPKSEDAGAGETPHMLLRPHVFMPEVTPDMOYLPRVFNQGIINPMLS 1099  
 DB 1081 QTPNSKLVENPKSEDAGAGETPHMLLRPHVFMPEVTPDMOYLPRVFNQGIINPMLS 1140  
 QY 1100 DETNICNGKPDVGLTTLRLNGTLVAFRGHYFWMLSPFSPSPARRITEVWGIPSIDTVFT 1159  
 DB 1141 DETNICNGKPDVGLTTLRLNGTLVAFRGHYFWMLSPFSPSPARRITEVWGIPSIDTVFT 1200  
 QY 1160 RCNCEGHTFFKDSQYWRFTNDIKDAGYKPIFKGFGGLTQGIVAALSTAKYKNWPSVY 1219  
 DB 1201 RCNCEGHTFFKDSQYWRFTNDIKDAGYKPIFKGFGGLTQGIVAALSTAKYKNWPSVY 1260  
 QY 1220 FEKGGSIQQYIYKQEPVQKCPGRPALNYPVYGVEMTQVRRRFEAIGPSQTHIRIQY 1279  
 DB 1261 FEKGGSIQQYIYKQEPVQKCPGRPALNYPVYGVEMTQVRRRFEAIGPSQTHIRIQY 1320  
 QY 1280 SPARLAYQDKGVHLNEVKVSIILWRGLPNVVTSAISLNNIRKPDGYDYAFSKDQYNNIDV 1339  
 DB 1321 SPARLAYQDKGVHLNEVKVSIILWRGLPNVVTSAISLNNIRKPDGYDYAFSKDQYNNIDV 1380  
 QY 1340 PSARAIAITRSGQTLSKWNVNC 1363  
 DB 1381 PSARAIAITRSGQTLSKWNVNC 1404

## RESULT 4

ADM98014

ID ADM98014 standard; protein; 1404 AA.

XX

AC ADM98014;

XX

DT 01-JUL-2004 (first entry)

XX

DE Human megakaryocyte stimulating factor (MSF).

XX

KW lubricating polypeptide; O-linked oligosaccharide; joint lubrication; CAP; camptodactyl-arthropathy pericarditis; osteoarthritis; human;

XX

KW megakaryocyte stimulating factor; MSF.

XX

OS Homo sapiens.

XX

PN US2004072741-A1.

XX

PD 15-APR-2004.

XX

PF 02-JUL-2001; 2001US-00897188.

XX

PR 23-APR-1999; 99US-00298970.

XX

PR 24-APR-2000; 2000US-00556246.

XX

PA (JAYC/) JAY G D.

XX

PI Jay GD;

XX

DR WPI; 2004-373948/35.

XX

DR N-PSDB; ADM98015.

XX

XX New tribonectin polypeptides and polynucleotides for lubricating joints  
 PT or other tissues to prevent or treat Camptodactyl-arthropathy-  
 PT pericarditis syndrome or osteoarthritis.  
 XX Claim 1; SEQ ID NO 1; 34pp; English.

XX

CC The invention relates to a lubricating polypeptide and at least one O-  
 CC linked oligosaccharide. The composition and methods are useful for  
 CC lubricating joints or other tissues to prevent or treat camptodactyl-  
 CC arthropathy pericarditis (CAP) or osteoarthritis in mammals. The present  
 CC sequence represents the amino acid sequence of the human megakaryocyte

CC

stimulating factor (MSF).

XX

SQ Sequence 1404 AA;

Query Match

Best Local Similarity 99.8%; Score 7261.9; DB 8; Length 1404;

Matches 1363; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY

1 MAWKTLPIYLLLSVFVIQVSSQ-----

DB 1 MAWKTLPIYLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNDYNCOHYMECCPDF 60

QY 26 -----ELSCGRCFESFERGREGCDCAQCKYDKCCPDYESFCAEVHNTPSPSSKKAP 79

DB 61 KRVCITABLSCKGRCFESFERGREGCDCAQCKYDKCCPDYESFCAEVHNTPSPSSKKAP 120

QY 80 PPSGASQTIKSTTKRSPKPNKKTKKVISEETIEHSVSENOESSSSSSSSSSSTI 139

DB 121 PPSGASQTIKSTTKRSPKPNKKTKKVISEETIEHSVSENOESSSSSSSSSSSTI 180

QY 140 KIKSSKNSAANRELQKKLVKDNKKNRKKKTPKPPVWDEAGSLDNGDFKVTTPDTST 199

DB 181 KIKSSKNSAANRELQKKLVKDNKKNRKKKTPKPPVWDEAGSLDNGDFKVTTPDTST 240

QY 200 TOHMKVSTSPKITTAKPINPRPSLPNSDTSKETSILTVMKETTIVTKETTTNKQSTDG 259

DB 241 TOHMKVSTSPKITTAKPINPRPSLPNSDTSKETSILTVMKETTIVTKETTTNKQSTDG 300

QY 260 KEKTTSAKETQSIEKTSADLAPTSKVLAKPTKAEITTKGPAITTKPEPTTTTKEP 319

DB 301 KEKTTSAKETQSIEKTSADLAPTSKVLAKPTKAEITTKGPAITTKPEPTTTTKEP 360

QY 320 TTPKEPTTTIKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 379

DB 361 TTPKEPTTTIKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 420

QY 380 APTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 439

DB 421 APTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 480

QY 440 EPAPTAPKAPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 499

DB 481 EPAPTAPKAPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 540

QY 500 TTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 559

DB 541 TTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 600

QY 560 APTAPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 619

DB 601 APTAPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 660

QY 620 PEEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 679

DB 661 PEEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 720

QY 680 APTTPKAPKELAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 739

DB 721 APTTPKAPKELAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 780

QY 740 TAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 799

DB 781 TAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 840

QY 800 KPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 859

DB 841 KPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 900

QY 860 TKTAAIKPEMTTAAKDKTTERDLRTTPTTAAAPKMTKETTATTEKTESKIATTTTQV 919

DB 901 TKTAAIKPEMTTAAKDKTTERDLRTTPTTAAAPKMTKETTATTEKTESKIATTTTQV 960

QY 920 TSTTTQDTTPPKIITLTKTTLLAPKVTTTKKTTTITTEIMNKPEETAAPKDRATNSKATTPK 979



Dd 792 TAPTTLKEPAPTTKKKAPKELAPTTTKGPTSTTSDDKPAPTTPKETAFTTPKEPAPTPK 851  
 Qy 800 KPAPTTTPTPTTSEVSTPTTTTKEPTTIHKSPDESTPELSAEPTEKALENSKPEPGVPT 859  
 Dd 852 KPAPTTTPTPTTSEVSTPTTTTKEPTTIHKSPDESTPELSAEPTEKALENSKPEPGVPT 911  
 Qy 860 TKTPAAATKPEMTTAKOKITRDLRTTPTTETTAAPKMTKETATTTTEKTESKITATTQV 919  
 Dd 912 TKTPAAATKPEMTTAKOKITRDLRTTPTTETTAAPKMTKETATTTTEKTESKITATTQV 971  
 Qy 920 TSTTTQDTPPKITTLTKTTLAPKVTTKKIIITTEIMNKPEETAAPKDRATNSKATTPK 979  
 Dd 972 TSTTTQDTPPKITTLTKITTLAPKVTTKKIIITTEIMNKPEETAAPKDRATNSKATTPK 1031  
 Qy 980 PQKPTAKPKPTSTKPKTWPVRVKPKTTPPRKWTSTMPENLPTSRIAEAMLOTTTRPN 1039  
 Dd 1032 PQKPTAKPKPTSTKPKTWPVRVKPKTTPPRKWTSTMPENLPTSRIAEAMLOTTTRPN 1091  
 Qy 1040 QTNNSKLIVEVNPKSSEDAGGAGETPHMLLRPHVFMPEVTDMOYLPRVPNQGIINPMLS 1099  
 Dd 1092 QTNNSKLIVEVNPKSSEDAGGAGETPHMLLRPHVFMPEVTDMOYLPRVPNQGIINPMLS 1151  
 Qy 1100 DETNICNGKXPVDGLTTLRNGTLVAFRGHYFWMLSPPSPSPARRITEVWGIPSPIDTVFT 1159  
 Dd 1152 DETNICNGKXPVDGLTTLRNGTLVAFRGHYFWMLSPPSPSPARRITEVWGIPSPIDTVFT 1211  
 Qy 1160 RCNCEGTFEFDKQSWRFNDIKDAGYKPIFKGFGGLTGQIVAAALSTAKYNWPEVY 1219  
 Dd 1212 RCNCEGTFEFDKQSWRFNDIKDAGYKPIFKGFGGLTGQIVAAALSTAKYNWPEVY 1271  
 Qy 1220 FFKRGSGIQYIYKQBPVQKCGRRPALNYPVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 1279  
 Dd 1272 FFKRGSGIQYIYKQBPVQKCGRRPALNYPVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 1331  
 Qy 1280 SPARLAYQDKGVLHNEKVUSILWGLPNVVTSAISLNPTRKPDGYDYAFSKDQYNNIDV 1339  
 Dd 1332 SPARLAYQDKGVLHNEKVUSILWGLPNVVTSAISLNPTRKPDGYDYAFSKDQYNNIDV 1391  
 Qy 1340 PSTARAITRSGQTLISKVWYNCP 1363  
 Dd 1392 PSTARAITRSGQTLISKVWYNCP 1415

RESULT 6  
 ADK65839  
 ID ADK65839 standard; protein; 1404 AA.  
 AC ADK65839;  
 XX  
 DT 06-MAY-2004 (first entry)  
 DE Angiogenesis-differentially expressed protein #53.  
 DE  
 DE cytotatic; cardiant; vasotropic; antiarteriosclerotic;  
 KW angiogenesis inhibitor; angiogenesis stimulator; angiogenic index;  
 KW gene expression; cancer; coronary artery disease; myocardial ischemia;  
 KW coronary arteriosclerosis; forensic medicine.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003066831-A2.  
 XX  
 PD 14-AUG-2003.  
 XX  
 PF 07-FEB-2003; 2003WO-US003848.  
 XX  
 PR 07-FEB-2002; 2002US-00067482.  
 PR 10-JUN-2002; 2002US-00164595.  
 PR 16-AUG-2002; 2002US-0403649P.  
 PR 03-JAN-2003; 2003US-0437746P.  
 XX  
 PA (ORIG-) ORIGENE TECHNOLOGIES INC.  
 XX

PI Sun Z, Li X, Kovacs KF, Fan W, Jay G;  
 XX WPI; 2003-731502/69.  
 DR  
 XX  
 PT Determining the angiogenic index of a tissue or cell sample using  
 PT expression levels of differentially expressed genes, useful for  
 PT diagnosing or treating cancer, coronary artery disease, myocardial  
 PT ischemia and/or arteriosclerosis.  
 XX  
 PS Disclosure; SEQ ID NO 78; 296pp; English.  
 XX  
 CC The invention relates to a method of determining the angiogenic index of  
 CC a tissue or cell sample comprising assessing, in a sample, the expression  
 CC levels of one or more differentially-expressed gene from any of 34 DNA  
 CC sequences, given in the specification, where the levels are indicative of  
 CC the angiogenic index. The methods and compositions of the present  
 CC invention are useful for diagnosing, preventing and/or treating cancer,  
 CC coronary artery disease, myocardial ischemia or coronary  
 CC arteriosclerosis. They can also be used in research, drug discovery and  
 CC forensic medicine involving angiogenesis. This sequence corresponds to  
 CC one of the differentially expressed proteins of the invention.  
 XX  
 SQ Sequence 1404 AA;

Query Match 99.4%; Score 7232.9; DB 7; Length 1404;  
 Best Local Similarity 96.8%; Pred. No. 4e-176;  
 Matches 1359; Conservative 0; Mismatches 4; Indels 41; Gaps 1;  
 Qy 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25  
 Dd 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYRDATCNCNDYQCNQHVMECCPDF 60  
 Qy 26 -----ELCKGRCFSPERGRECDCAQCKYDKCPDYESFCAEVHNPTSPSSKKAP 79  
 Dd 61 KRVTAEELCKGRCFSPERGRECDCAQCKYDKCPDYESFCAEVHNPTSPSSKKAP 120  
 Qy 80 PRSGASQTIKSTTKRSPKPNKKTKVIESEBITTEHSVSENOESSSSSSSSSTI 139  
 Dd 121 PRSGASQTIKSTTKRSPKPNKKTKVIESEBITTEHSVSENOESSSSSSSSSTIR 180  
 Qy 140 KIKSSKNSAANRELQKLVKONKNRTKKKPTPKPPVVDGAGSLDNGDFKVTITDST 139  
 Dd 181 KIKSSKNSAANRELQKLVKONKNRTKKKPTPKPPVVDGAGSLDNGDFKVTITDST 240  
 Qy 200 TQHNKVSTSPKITTAKPINPRPSLPNSDTSKETSITVNKEITVETKETTINKQSTDG 259  
 Dd 241 TQHNKVSTSPKITTAKPINPRPSLPNSDTSKETSITVNKEITVETKETTINKQSTDG 300  
 Qy 260 KEKTTSAKETOSIEKTSADLAPTSKVLAKPTPKAETTTKGPAITTPKEPTTPKEPAS 319  
 Dd 301 KEKTTSAKETOSIEKTSADLAPTSKVLAKPTPKAETTTKGPAITTPKEPTTPKEPAS 360  
 Qy 320 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTPKEPAPTTTKKEPAPTTTPKEPTTPKE 379  
 Dd 361 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTPKEPAPTTTKKEPAPTTTPKEPTTPKE 420  
 Qy 380 APTTTKSAPTPKEPAPTTPKPAPTTTPKEPAPTTTPKEPTTPTPKEPAPTTTPKEPTPK 439  
 Dd 421 APTTTKSAPTPKEPAPTTPKPAPTTTPKEPAPTTTPKEPTTPTPKEPAPTTTPKEPTPK 480  
 Qy 440 EPAPTAPKKPAPTTTPKEPAPTTTPKEPAPTTTPKEPTTPTPKEPAPTTTPKEPTPK 499  
 Dd 481 EPAPTAPKKPAPTTTPKEPAPTTTPKEPAPTTTPKEPTTPTPKEPAPTTTPKEPTPK 540  
 Qy 500 TTKSAPTTTPKEPTTPTPKEPAPTTTPKEPAPTTTPKEPTTPTPKEPAPTTTPKEPTPK 559  
 Dd 541 TTKSAPTTTPKEPTTPTPKEPAPTTTPKEPAPTTTPKEPTTPTPKEPAPTTTPKEPTPK 600  
 Qy 560 APTAPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPTTPTPKEPAPTTTPKEPTPK 619  
 Dd 601 APTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPTTPTPKEPAPTTTPKEPTPK 660  
 Qy 620 PEEPAPITPKAAAPNTPKPAPTTTPKEPAPTTTPKEPAPTTTPKEPTTPTPKEPTPK 679

Db	661	PEEPAPTTKAAAPNPKBPAPTTKPEAPATTPKPEAPATTPKETAATTTGKTAPTTLKEP	720
Qy	680	APTTPKAPKAPKELAPTTTKEPTSTTSDKAPATTPKGTAPTTKPEAPATTPKGP	739
Db	721	APTTPKAPKAPKELAPTTTKEPTSTTCDKAPATTPKGTAPTTKPEAPATTPKGP	780
Qy	740	TAPTTLKEPAPTTKAPKAPKELAPTTTKEPTSTTSDKAPATTPKETAATTPKPEAPATTPK	799
Db	781	TAPTTLKEPAPTTKAPKAPKELAPTTTKEPTSTTSDKAPATTPKETAATTPKPEAPATTPK	840
Qy	800	KAPATTPPETPPPTSVSTPTTKEPTTHKSPDESTPELSAPTPKALENSKEGVP	859
Db	841	KAPATTPPETPPPTSVSTPTTKEPTTHKSPDESTPELSAPTPKALENSKEGVP	900
Qy	860	TKTPAATKPEMTTAKDKTTERDLRTTPETTTAAKMTKETAATTTKTESKITATTQV	919
Db	901	TKTPAATKPEMTTAKDKTTERDLRTTPETTTAAKMTKETAATTTKTESKITATTQV	960
Qy	920	TSSTTQDTPPKITTLKTTLAPKVTTTKTITTTIMNKPEETAKPKORATNSKATTPK	979
Db	961	TSSTTQDTPPKITTLKTTLAPKVTTTKTITTTIMNKPEETAKPKORATNSKATTPK	1020
Qy	980	POKETKAPKPTSTKKPKTMRVRKPKTTPTRKMTSTMPDELNPTSRIAEAMLOTTTRPN	1039
Db	1021	POKETKAPKPTSTKKPKTMRVRKPKTTPTRKMTSTMPDELNPTSRIAEAMLOTTTRPN	1080
Qy	1040	QTENSKLVEVNPXSSEDAGGEGEPHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS	1099
Db	1081	QTENSKLVEVNPXSSEDAGGEGEPHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS	1140
Qy	1100	DETNICNGKVPDGLTTLRNGTLVAFRGHYFWMLSPSPSPARRITEVWGIPSPIDTVFT	1159
Db	1141	DETNICNGKVPDGLTTLRNGTLVAFRGHYFWMLSPSPSPARRITEVWGIPSPIDTVFT	1200
Qy	1160	RCNCEGKTFEFKDSQYWRFTNDIKDAGYKPIFKGFGGLGQIIVAALSTAKYKNWPESVY	1219
Db	1201	RCNCEGKTFEFKDSQYWRFTNDIKDAGYKPIFKGFGGLGQIIVAALSTAKYKNWPESVY	1260
Qy	1220	FFKRGSGTQIYYKQEPVQKCPGRRPALNYPVYGETTQVRRRFRFAIGPSQTHTRIQY	1279
Db	1261	FFKRGSGTQIYYKQEPVQKCPGRRPALNYPVYGETTQVRRRFRFAIGPSQTHTRIQY	1320
Qy	1280	SPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLNPDKDGYDYAFSKQOYYNIDV	1339
Db	1321	SPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLNPDKDGYDYAFSKQOYYNIDV	1380
Qy	1340	PSRTARAITTRESGOTLSKVWYNCP	1363
Db	1381	PSRTARAITTRESGOTLSKVWYNCP	1404
RESULT 7			
ID	ADK65819	standard; protein; 1320 AA.	
XX	AC	ADK65819;	
XX	DT	06-MAY-2004 (first entry)	
XX	DE	Angiogenesis-differentially expressed protein ANH0316.	
XX	KW	cytostatic; cardiant; vasotropic; antiarteriosclerotic;	
XX	KW	angiogenesis inhibitor; angiogenesis stimulator; angiogenic index;	
XX	KW	gene expression; cancer; coronary artery disease; myocardial ischemia;	
XX	OS	coronary arteriosclerosis; forensic medicine.	
XX	PN	Homo sapiens.	
XX	PD	WO2003066831-A2.	
XX	PD	14-AUG-2003.	

PF	07-FEB-2003; 2003WO-US003848.		
XX	07-FEB-2002; 2002US-00067482.		
PR	10-JUN-2002; 2002US-00164595.		
PR	16-AUG-2002; 2002US-0403649P.		
PR	03-JAN-2003; 2003US-0437746P.		
XX	(ORIG-) ORIGENE TECHNOLOGIES INC.		
PA	Sun Z, Li X, Kovacs KF, Fan W, Jay G;		
XX	WPI; 2003-731502/69.		
PI	N-PSDB; ADK65818.		
DR	Determining the angiogenic index of a tissue or cell sample using		
XX	expression levels of differentially expressed genes, useful for		
PT	diagnosing or treating cancer, coronary artery disease, myocardial		
PT	ischemia and/or arteriosclerosis.		
XX	Claim 23; SEQ ID NO 58; 296pp; English.		
PS	The invention relates to a method of determining the angiogenic index of		
XX	a tissue or cell sample comprising assessing, in a sample, the expression		
CC	levels of one or more differentially-expressed gene from any of 34 DNA		
CC	sequences, given in the specification, where the levels are indicative of		
CC	the angiogenic index. The methods and compositions of the present		
CC	invention are useful for diagnosing, preventing and/or treating cancer,		
CC	coronary artery disease, myocardial ischemia or coronary		
CC	arteriosclerosis. They can also be used in research, drug discovery and		
CC	forensic medicine involving angiogenesis. This sequence corresponds to		
CC	one of the differentially expressed proteins of the invention.		
XX	Sequence 1320 AA;		
Qy	Query Match	96.8%; Score 7043.7; DB 7; Length 1320;	
Db	Best Local Similarity	96.6%; Pred. No. 2.5e-171;	
Qy	Matches 1317; Conservative	0; Mismatches 3; Indels 43; Gaps 1;	
Db	1	MAWKTPILYLLLLSVFVIOQVSSQELSCKGRFCFESPERGECDCDAQCKKYDKCCPDYE	60
Qy	61	SFCAEVHNPTSPSSKKAPPPSGASQITKTTKRSPPNKKTKKVIIESEITEHSVS	120
Db	61	SFCAEVHNPTSPSSKKAPPPSGASQITKTTKRSPPNKKTKKVIIESEITE-----	115
Qy	121	ENQESSSSSSSSSTIWIKSSKNSAANRELOKLLKVKDKNKNRTKKKTPKPPVVDE	180
Db	116	-----VKDNKNRTKKKTPKPPVVDE	137
Qy	181	AGSLDNGDFKVTTPDSTTQHNKVSTSPKITTAKPINPRPSPSPNSDTSKETSITVNKE	240
Db	138	AGSLDNGDFKVTTPDSTTQHNKVSTSPKITTAKPINPRPSPSPNSDTSKETSITVNKE	197
Qy	241	TTVETKETTNTKQSTDGKETTSAKETOSIEKTSAKOLAPTQKVLAKPTPKAETTTKG	300
Db	198	TTVETKETTNTKQSTDGKETTSAKETOSIEKTSAKOLAPTQKVLAKPTPKAETTTKG	257
Qy	301	PALTTPKEPTTPKPEASTTPKEPTTTIKSAPTTPKEAPATTTKSAPTTPKEAPATTT	360
Db	258	PALTTPKEPTTPKPEASTTPKEPTTTIKSAPTTPKEAPATTTKSAPTTPKEAPATTT	317
Qy	361	KEPAPTTKPEAPATTTKPEAPATTTKSAPTTPKPEAPATTTKSAPTTPKPEPTT	420
Db	318	KEPAPTTKPEAPATTTKPEAPATTTKSAPTTPKPEAPATTTKSAPTTPKPEPTT	377
Qy	421	TPKPEAPTTKPEAPTTAPKPPADTTTPKBPATTTKPEAPATTTKPEPTTKE	480
Db	378	TPKPEAPTTKPEAPTTAPKPPADTTTPKBPATTTKPEAPATTTKPEPTTKE	437
Qy	481	PAPTTTKSAPTTKPEAPTTKSAPTTPKPSPTTTKPEAPTTTPKPEAPTTPKKAPATTP	540
Db	438	PAPTTTKSAPTTKPEAPTTKSAPTTPKPSPTTTKPEAPTTTPKPEAPTTPKKAPATTP	497



QY 181 AGSGLDNGDFKVTTPDTSTTOHKNVSTSPKITTAKENINPRPSLPNSDTSKETSIVNKE 240  
DB 88 AGSGLDNGDFKVTTPDTSTTOHKNVSTSPKITTAKENINPRPSLPNSDTSKETSIVNKE 147  
QY 241 TTVETKETTNNKQSTDCGKETTSAKETQSTSEKTSKADLAPTSKVLAKPTPKAETTTKG 300  
DB 148 TTVETKETTNNKQSTDCGKETTSAKETQSTSEKTSKADLAPTSKVLAKPTPKAETTTKG 207  
QY 301 PALTTTKEPTPTTPKPEASTTPKEPTTTIKSAPTTPKGPAPTTTKSAPTTPKGPAPTTT 360  
DB 208 PALTTTKEPTPTTPKPEASTTPKEPTTTIKSAPTTPKGPAPTTTKSAPTTPKGPAPTTT 267  
QY 361 KEPAPTTPKEPAPTTTKSAPTTPKAPTTTPKAPTTTPKAPTTTPKAPTTTPKAPTTTP 420  
DB 268 KEPAPTTPKEPAPTTTKSAPTTPKAPTTTPKAPTTTPKAPTTTPKAPTTTPKAPTTTP 327  
QY 421 TTPKEPAPTTPKEPAPTTPKAPTTTPKAPTTTPKAPTTTPKAPTTTPKAPTTTPKAPTTTP 480  
DB 328 TTPKEPAPTTPKEPAPTTPKAPTTTPKAPTTTPKAPTTTPKAPTTTPKAPTTTPKAPTTTP 387  
QY 481 PAPTTPKSAPTTTKSAPTTPKAPTTTPKAPTTTPKAPTTTPKAPTTTPKAPTTTPKAPTTTP 540  
DB 388 PAPTTPKSAPTTTKSAPTTPKAPTTTPKAPTTTPKAPTTTPKAPTTTPKAPTTTPKAPTTTP 447  
QY 541 KEPAPTTPKEPAPTTPKAPTTTPKAPTTTPKAPTTTPKAPTTTPKAPTTTPKAPTTTPKAPTTTP 600  
DB 448 KEPAPTTPKEPAPTTPKAPTTTPKAPTTTPKAPTTTPKAPTTTPKAPTTTPKAPTTTPKAPTTTP 507  
QY 601 PTPPEELAPTTPEEPTPTTPPEEPAPTTPKAAAPNTPKAPTTTPKAPTTTPKAPTTTPKAPTTTP 660  
DB 508 PTPPEELAPTTPEEPTPTTPPEEPAPTTPKAAAPNTPKAPTTTPKAPTTTPKAPTTTPKAPTTTP 567  
QY 661 KETAPTTPKGTAPTTLUKEPAPTTPKAPKELAPTTTKEPTSTTSKAPAPTTTPKAGPAPT 720  
DB 568 KETAPTTPKGTAPTTLUKEPAPTTPKAPKELAPTTTKEPTSTTSKAPAPTTTPKAGPAPT 627  
QY 721 PKAPAPTTPKAPAPTTPKGTAPTTLUKEPAPTTPKAPKELAPTTTKEPTSTTSKAPAPTTTP 780  
DB 628 PKAPAPTTPKAPAPTTPKGTAPTTLUKEPAPTTPKAPKELAPTTTKEPTSTTSKAPAPTTTP 687  
QY 781 TPKEPATTPKEPATTPPKAPAPTTTPPTTPPTTSEVSTPTTKEPTTIHKSPDESTPELS 840  
DB 688 TPKEPATTPKEPATTPPKAPAPTTTPPTTPPTTSEVSTPTTKEPTTIHKSPDESTPELS 747  
QY 841 AEPTPKALENSPKBPGVPTTKPAATKPEMTTAKOKTTERDLRTTETTTAAPKMTKET 900  
DB 748 AEPTPKALENSPKBPGVPTTKPAATKPEMTTAKOKTTERDLRTTETTTAAPKMTKET 807  
QY 901 ATTTTEKTESKITATTQTSTTQDTPPKITLTKTTLAPKVTITTKITTTTTEIMNKP 960  
DB 808 ATTTTEKTESKITATTQTSTTQDTPPKITLTKTTLAPKVTITTKITTTTTEIMNKP 867  
QY 961 EETAAPKDRATNSKATTPKQKPTKAPKPTSTKPKPTMPVRKPTTTPRKMTSTWPE 1020  
DB 868 EETAAPKDRATNSKATTPKQKPTKAPKPTSTKPKPTMPVRKPTTTPRKMTSTWPE 927  
QY 1021 LNPTSRJAEAMLQTTTRPNQTPNSKLVNPKSEDAGAGETPHMLLRPHVFPVEVTPD 1080  
DB 928 LNPTSRJAEAMLQTTTRPNQTPNSKLVNPKSEDAGAGETPHMLLRPHVFPVEVTPD 987  
QY 1081 MDYLPRVNOGIIINPMLSDETNICNGKPVGLTTLRNGTLVAFRGHYFWMLSFSPSP 1140  
DB 988 MDYLPRVNOGIIINPMLSDETNICNGKPVGLTTLRNGTLVAFRGHYFWMLSFSPSP 1047  
QY 1141 ARRITEVWGIPSIDTVTRCNCEGKTFPFDKDSQVWRFTNDIKDAGYKPKTFKFGGLGT 1200  
DB 1048 ARRITEVWGIPSIDTVTRCNCEGKTFPFDKDSQVWRFTNDIKDAGYKPKTFKFGGLGT 1107  
QY 1201 QIVAAALSTAKVNPESVYFKRGSIQOQYIKQEPVQKCGRRPALNYPVYGETQVRR 1260  
DB 1108 QIVAAALSTAKVNPESVYFKRGSIQOQYIKQEPVQKCGRRPALNYPVYGETQVRR 1167  
QY 1261 RFRERAIGPSQTHTIRIQSPARLAYQDKGVILHNEVKVSIILWRGLPNVVTSAISLPIR 1320

DB 1168 RFRERAIGPSQTHTIRIQSPARLAYQDKGVILHNEVKVSIILWRGLPNVVTSAISLPIR 1227  
QY 1321 PDGYDYAIFSKDQYNNIDVPSRTARAITTRSGQTLSKVMYNCP 1363  
DB 1228 PDGYDYAIFSKDQYNNIDVPSRTARAITTRSGQTLSKVMYNCP 1270  
RESULT 9  
ADK67912  
ID ADK67912 standard; protein; 1311 AA.  
XX  
AC ADK67912;  
XX  
DT 06-MAY-2004 (first entry)  
XX  
DE Human extracellular messenger (EXMES) polypeptide.  
XX  
KW Human; extracellular messenger; EXMES; respiratory-gen.; anti-allergic;  
KW antiasthmatic; anti-inflammatory; antidiabetic; neuroprotective;  
KW muscular-gen.; antiarthritic; osteopathic; hepatotropic; antipsoriatic;  
KW varicide; fungicide; antiparasitic; protozoacide; antihelminthic;  
KW cytostatic; gene therapy.  
XX  
OS Homo sapiens.  
XX  
PH Key Location/Qualifiers  
FT Peptide 1..30 /label= Signal peptide  
FT /note= "Spans residues 1 to 18, 20, 21, 24, 29 or 30 according to identification method"  
XX  
PN WO2004013292-A2.  
XX  
PD 12-FEB-2004.  
XX  
PF 30-JUL-2003; 2003WO-US024084.  
XX  
PR 02-AUG-2002; 2002US-0400810P.  
PR 19-SEP-2002; 2002US-0412197P.  
PR 04-OCT-2002; 2002US-0416004P.  
PR 08-NOV-2002; 2002US-0424862P.  
XX  
PA (INCY-) INCYTE CORP.  
XX  
PI Elliott VS, Khare R, Tran UK, Swarnakar A, Marquis JP;  
PI Richardson TW, Emerling BM, Lindquist EA, Chawla NK, Ramkumar J;  
PI Lee SY;  
XX  
WPI: 2004-157116/15.  
N-PSDB; ADK67917.  
XX  
New extracellular messengers and nucleic acids, useful for diagnosing, treating or preventing e.g. allergies, anemia, Crohn's disease, diabetes mellitus, myasthenia gravis, osteoporosis, psoriasis, cancers or autoimmune thyroiditis.  
XX  
PS Claim 60; SEQ ID NO 5; 165pp; English.  
XX  
The present sequence is that of novel human extracellular messenger (EXMES) incyte ID NO: 7513018CD1 polypeptide. The protein shows homology to human megakaryocyte stimulating factor. The invention provides EXMES polynucleotides and polypeptides, as well as expression vectors, host cells, antibodies, agonists and antagonists, and methods for diagnosing, treating or preventing disorders associated with aberrant expression of EXMES, especially autoimmune and inflammatory disorders, cell proliferative disorders and endocrine disorders, e.g. adult respiratory distress syndrome, allergy, anaemia, erythroblastosis foetalis, Crohn's disease, diabetes mellitus, myasthenia gravis, osteoarthritis, osteoporosis, pancreatitis, psoriasis, viral, bacterial, fungal, parasitic, protozoal or helminthic infections, cancers, autoimmune thyroiditis, cretinism, Plummer's disease or thyroid carcinoma. Embodiments also provide methods for using the purified EXMES and/or

CC their encoding polynucleotides for facilitating the drug discovery  
CC process, including determining of efficacy, dosage, toxicity and  
CC pharmacology, and for investigating the pathogenesis of diseases and  
CC medical conditions.

SQ Sequence 1311 AA;

Query Match 93.1%; Score 6774.6; DB 8; Length 1311;  
Best Local Similarity 90.4%; Pred. No. 1.8e-164;  
Matches 1269; Conservative 0; Mismatches 1; Indels 134; Gaps 2;

QY	1	MAMKTLPIYLILLLLSVFVIQOVSSQ	-----	25
Db	1	MAMKTLPIYLILLLLSVFVIQOVSSQDLSSACRGCGEYSRDATCNCDVNCQHYMECCPDF	60	
QY	26	-----ELSCKGRCFESFERGECDDAOCKKYDKCCPDYBSFCAEVHNTSPSPSSKAP	79	
Db	61	KRVCTAELSCKGRCFESFERGECDDAOCKKYDKCCPDYBSFCAE	-----	106
QY	80	PPGASQTIKSTTKRSPKPPNKKTKKVIIESEBITEEHSVSENQESSSSSSSSSTI	139	
Db	107	-----	-----	106
QY	140	KIKSSKNSAANRELQKKLVKONKKNRTKKKPTPKPPVWDEAGSLDNGDFKVTTPDTST	199	
Db	107	-----VKDNKKNRTKKKPTPKPPVWDEAGSLDNGDFKVTTPDTST	147	
QY	200	TOHNKYSTPKITTAKEPINRPSLPNDSDTSKETSLTVNKETTVETKETTNNKQTS	259	
Db	148	TOHNKYSTPKITTAKEPINRPSLPNDSDTSKETSLTVNKETTVETKETTNNKQTS	207	
QY	260	KEKTTSAKETQSIKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPEKPTPTPKEPAS	319	
Db	208	KEKTTSAKETQSIKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPEKPTPTPKEPAS	267	
QY	320	TTPEKPTPTTIKSAPTTPKEPAPTTTKSAPTTPEKAPTTTKGPAPTTPEKAPTTTKEP	379	
Db	268	TTPEKPTPTTIKSAPTTPKEPAPTTTKSAPTTPEKAPTTTKGPAPTTTKEP	327	
QY	380	APTPTTKSAPTTPEKAPTTTPKKAPPTTPKEPAPTTTPKEPPTTPPKGPAPTTKGPAPTTPK	439	
Db	328	APTPTTKSAPTTPEKAPTTTPKKAPPTTPKEPAPTTTPKEPPTTPPKGPAPTTKGPAPTTPK	387	
QY	440	EPAPTAPKKAPPTTPKEPAPTTPEKAPTTTKEPSPTTPKEPAPTTTKSAPTTTKGPAPT	499	
Db	388	EPAPTAPKKAPPTTPKEPAPTTTPKEPAPTTTKEPSPTTPKEPAPTTTKSAPTTTKGPAPT	447	
QY	500	TTKSAPTTTKEPSPTTKGPAPTTTPKEPAPTTTPKKAPPTTPKEPAPTTTPKEPAPTTTKKP	559	
Db	448	TTKSAPTTTKEPSPTTKGPAPTTTPKEPAPTTTPKKAPPTTPKEPAPTTTPKEPAPTTTKKP	507	
QY	560	ADTAPKEPAPTTPKETAPTTPKLTPPTPEKLAPTTPEKAPTTTPSELAPTTPEEPPTPT	619	
Db	508	ADTAPKEPAPTTPKETAPTTPKLTPPTPEKLAPTTPEKAPTTTPSELAPTTPEEPPTPT	567	
QY	620	PPEPAPTTPKAAAPNTPKEPAPTTTPKEPAPTTTPKEPAPTTPKETADTTPKGTAPTTLKBP	679	
Db	568	PPEPAPTTPKAAAPNTPKEPAPTTTPKEPAPTTTPKEPAPTTPKETADTTPKGTAPTTLKBP	627	
QY	680	APTTPKKAPKELAPTTTKETSTTSDKAPTTPKGTAPTTPEKAPTTTPKEPAPTTPKGP	739	
Db	628	APTTPKKAPKELAPTTTKETSTTSDKAPTTTPKGATTTTPKEPAPTTTPKGPAPTTPKG	687	
QY	740	TAPTTLKEPAPTTPKKAPKELAPTTTKGTSTTSDKAPTTTPKETAPTTTPKEPAPTTPK	799	
Db	688	TAPTTLKEPAPTTPKKAPKELAPTTTKGTSTTSDKAPTTTPKETAPTTTPKEPAPTTPK	747	
QY	800	KZAPTTTPETPTTSEVSTPTTKGPPTTIHKSPDSESTPELSAEPPTKALENSFKPFGVPT	859	
Db	748	KZAPTTTPETPTTSEVSTPTTKGPPTTIHKSPDSESTPELSAEPPTKALENSFKPFGVPT	807	
QY	860	TKTTPAATKPEMUTTKAKDKTTERDLATTPETTTTAAAPKMTKETATTTKTESKIATTTQV	919	

Db	808	TKTPAATKPEMTTAKOKXITRDLRTTPEITTTAAAPKMTKETATTTTEKTESKITATTQV	867
QY	920	TSITTTQDITTPFKITITLTKITTIAPKVTTTKTITTEIMNKPEETAAPKDRATNSKATTPK	979
Db	868	TSITTTQDITTPFKITITLTKITTIAPKVTTTKTITTEIMNKPEETAAPKDRATNSKATTPK	927
QY	980	FOKPTKAPKPKSTSKKPKTMPRVKPKTTPTPRKMTSTMPELNPTSRIAEAMIQTTTRPN	1039
Db	928	POKPTKAPKPKSTSKKPKTMPRVKPKTTPTPRKMTSTMPELNPTSRIAEAMIQTTTRPN	987
QY	1040	QTPNSKLVEVNPKSEDAGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIINPMLS	1099
Db	988	QTPNSKLVEVNPKSEDAGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIINPMLS	1047
QY	1100	DETNIKCGKPDVGLTTLRNGTLVAFRGHYFWMLSPFSPSPARRITEVWGISPIDTFT	1159
Db	1048	DETNIKCGKPDVGLTTLRNGTLVAFRGHYFWMLSPFSPSPARRITEVWGISPIDTFT	1107
QY	1160	RCNCEGKTFFPKDSQYWRFTNDIKDAGYKPIFKFGGLTGQIVAAALSTAKYKNWPESVY	1219
Db	1108	RCNCEGKTFFPKDSQYWRFTNDIKDAGYKPIFKFGGLTGQIVAAALSTAKYKNWPESVY	1167
QY	1220	FFKGGSIQQYIYKQEPVQKCPGRRPALNYPVYGEVTQVRRRRFERAIGPSQHTTIRIQY	1279
Db	1168	FFKGGSIQQYIYKQEPVQKCPGRRPALNYPVYGEVTQVRRRRFERAIGPSQHTTIRIQY	1227
QY	1280	SPARLAIQDKGVLHNEVKVLSILWRGLPNVVTSAISLPIKRPDGYDYAFSKDQYYNIDV	1339
Db	1228	SPARLAIQDKGVLHNEVKVLSILWRGLPNVVTSAISLPIKRPDGYDYAFSKDQYYNIDV	1287
QY	1340	PSRTARAITTRSGOTLSKWYNCP	1363
Db	1288	PSRTARAITTRSGOTLSKWYNCP	1311
RESULT	10		
AAM24322			
ID	AAM24322	standard; protein; 1299 AA.	
AC	AAM24322;		
XX			
DT	12-OCT-2001	(first entry)	
XX			
DE	Human EST	encoded protein SEQ ID NO: 1847.	
DE			
DE			
KW	Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;		
KW	tomato; monkey; dog; sea urchin; expressed sequence tag; EST;		
KW	diagnostics; forensic test; gene mapping; genetic disorder; biodiversity;		
KW	gene therapy; nutrition.		
XX			
OS	Homo sapiens.		
XX			
EN	WO200154477-A2.		
XX			
PD	02-AUG-2001.		
XX			
PF	25-JAN-2001; 2001WO-US002687.		
XX			
PR	25-JAN-2000; 2000US-00491404.		
PR	17-JUL-2000; 2000US-00617746.		
PR	03-AUG-2000; 2000US-00631451.		
PR	15-SEP-2000; 2000US-00663870.		
XX			
PA	(HYSE-) HYSEQ INC.		
XX			
PI	Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;		
PI	Cao Y, Drmanac RA, Zhang J, Werhman T;		
XX			
DR	WPI; 2001-476164/51.		
XX			
XX	N-PSDB; AAH98981.		
PT	Isolated polypeptide for treatment of diseases, diagnostics, raising		
PT	antibodies and research use		



XX PS Claim 20; Page 1198-1201; 1275pp; English.

XX CC The present invention provides the protein and coding sequences of novel

CC proteins from a variety of organisms, including human, dog, cat, horse,

CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea

CC urchin and tomato. These were derived from expressed sequence tags (ESTs)

CC from the organism of interest. They can be used in diagnostics,

CC forensics, gene mapping, identification of mutations, to assess

CC biodiversity and for nutritional purposes. The present sequence is a

CC protein of the invention

XX SQ Sequence 1299 AA;

Query Match 91.9%; Score 6685.9; DB 4; Length 1299;

Best Local Similarity 96.8%; Pred. No. 3.3e-162;

Matches 1254; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAWKTLPIYILLLLSVFVIQQVSSQ----- 25

DB 1 MAWKTLPIYILLLLSVFVIQQVSSQ----- 60

QY 26 -----ELCKGRCEFSFREGRCDDAQCCKYDKCCPDYESFCAEVHNPTSPSSKKAP 79

DB 61 KRVTALSCCKGRCEFSFREGRCDDAQCCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120

QY 80 PPSGASQTKSTTKRSPKPPNKKTKKVIIESEITEHSHSVSEHQESSSSSSSSSSSTIW 139

DB 121 PPSGASQTKSTTKRSPKPPNKKTKKVIIESEITEHSHSVSEHQESSSSSSSSSSSTIW 180

QY 140 KIKSSKNSAANRELQKLVKDNKKNTKKPKPPVDEAGSLDNGDFKVTTPDTST 199

DB 181 KIKSSKNSAANRELQKLVKDNKKNTKKPKPPVDEAGSLDNGDFKVTTPDTST 240

QY 200 TQNKVSTSPKITTAKPINRPSLPNPSDTSKETSIVNKEITVETKETTINKQSTGD 259

DB 241 TQNKVSTSPKITTAKPINRPSLPNPSDTSKETSIVNKEITVETKETTINKQSTGD 300

QY 260 KEKTSAKETQSIEKTSKOLAPTSKVLAKPTPKAETTTKGPALTTKPEPTPTTKBPAS 319

DB 301 KEKTSAKETQSIEKTSKOLAPTSKVLAKPTPKAETTTKGPALTTKPEPTPTTKBPAS 360

QY 320 TTPKEPTPTTKSAPTTKBPAPTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBP 379

DB 361 TTPKEPTPTTKSAPTTKBPAPTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBP 420

QY 380 APTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTK 439

DB 421 APTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTK 480

QY 440 EPAPTAPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 499

DB 481 EPAPTAPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 540

QY 500 TTKSAPTTKBPSPPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTK 559

DB 541 TTKSAPTTKBPSPPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTK 600

QY 560 APTAPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 619

DB 601 APTAPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 660

QY 620 PEEPAPTTKAAAPNTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 679

DB 661 PEEPAPTTKAAAPNTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 720

QY 680 APTPKKBPAPKELAPTTKBPSTSDKBPAPTTKBPAPTTKBPAPTTKBPAPTT 739

DB 721 APTPKKBPAPKELAPTTKBPSTSDKBPAPTTKBPAPTTKBPAPTTKBPAPTT 780

QY 740 TAPTTLKBPAPTTKBPAPKELAPTTKGTSTSDKBPAPTTKBPAPTTKBPAPTT 799

DB 781 TAPTTLKBPAPTTKBPAPKELAPTTKGTSTSDKBPAPTTKBPAPTTKBPAPTT 840

QY 800 KPAPTTPEPTTSEVSTPTTTKBPPTTIHKSPDESTPELSAEPPTKALENSPKBPGVPT 859

DB 841 KPAPTTPEPTTSEVSTPTTTKBPPTTIHKSPDESTPELSAEPPTKALENSPKBPGVPT 900

QY 860 TKTPAATKPEMTTAAKDKTTERDLRTTPTTTAAPKOTKETATTTTETTESKITATTQV 919

DB 901 TKTPAATKPEMTTAAKDKTTERDLRTTPTTTAAPKOTKETATTTTETTESKITATTQV 960

QY 920 TSTTTQDTPPKITTLKTTTLAPKVTITTKITTTTTEIMNKPETAKPKORATNSKATTPK 979

DB 961 TSTTTQDTPPKITTLKTTTLAPKVTITTKITTTTTEIMNKPETAKPKORATNSKATTPK 1020

QY 980 POKPTKAPKPTSTKPKTMPRVKPKTPTTPKMTSTMPELNPTSRIAEAMLQTTTRPN 1039

DB 1021 POKPTKAPKPTSTKPKTMPRVKPKTPTTPKMTSTMPELNPTSRIAEAMLQTTTRPN 1080

QY 1040 QTPNSKLVEVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMYLRVPVNPQGIINPMLS 1099

DB 1081 QTPNSKLVEVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMYLRVPVNPQGIINPMLS 1140

QY 1100 DETNICNGKPVVDGLTTLRNGTLVAFRGHYFWMLSPEFSPSPARRITEVWGIPIPIDIVFT 1159

DB 1141 DETNICNGKPVVDGLTTLRNGTLVAFRGHYFWMLSPEFSPSPARRITEVWGIPIPIDIVFT 1200

QY 1160 RCNCEGKTRFFKDSQYWRFTNDIKDAGYKPIFKFGGLTSGQIVAAALSTAKYKNWPESVY 1219

DB 1201 RCNCEGKTRFFKDSQYWRFTNDIKDAGYKPIFKFGGLTSGQIVAAALSTAKYKNWPESVY 1260

QY 1220 FFKRGGSIQYIYKQBPVQKCPGRRPALNYPVYGE 1254

DB 1261 FFKRGGSIQYIYKQBPVQKCPGRRPALNYPVYGE 1295

RESULT 11

AAB29778

ID AAB29778 standard; protein; 902 AA.

XX AC AAB29778;

XX DT 28-FEB-2001 (first entry)

XX DE Human MSF-derived tribonectin.

XX KM Human tribonectin; MSF, megakaryocyte stimulating factor;

XX KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;

XX KW osteoarthritis; tribosupplementation; tissue adhesion inhibition;

XX KW friction coefficient reduction; gene therapy; antiarthritic; osteopathic.

XX OS Homo sapiens.

XX FN WO200064930-A2.

XX PD 02-NOV-2000.

XX PF 24-APR-2000; 2000WO-US010953.

XX PR 23-APR-1999; 99US-00298970.

XX PA (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.

XX PI Jay GD;

XX DR WPI; 2001-024673/03.

XX PT Novel tribonectin polypeptide useful as lubricant for treating

XX PT osteoarthritis, comprises O-linked lubricating moiety.

XX PS Disclosure; Fig 1; 47pp; English.

XX CC The invention relates to a human tribonectin which is a product of

CC CC alternative splicing of the human MSF (megakaryocyte stimulating factor)

CC CC gene. The tribonectin has at least one O-linked oligosaccharide



lubricating moiety and has a polypeptide sequence comprising 1-76 repeats of a motif having at least 50% identity to the sequence KEPAPTT (AAB23774). The invention also relates to a nucleic acid encoding a human MSF-derived tribonection; a biocompatible composition comprising a human tribonection for inhibiting tissue adhesion formation; and a method of diagnosing osteoarthritis or a predisposition to osteoarthritis by measuring the amount of MSF or its fragment in a biological sample of a mammal, wherein an increased amount of MSF compared to a control indicates the presence of or predisposition to developing osteoarthritis. The tribonection and DNA encoding it are useful in the treatment of osteoarthritis, where they may be used for lubricating mammalian joints, such as articulating joints of humans, dogs or horses. The tribonection, when formulated as a membrane, foam, gel or fibre, is useful for inhibiting adhesion between two surfaces such as the injured tissues of a mammal, where the injury is caused by a surgical insertion or trauma, or an artificial device e.g., an orthopaedic implant. In particular, one of the surfaces is pericardial tissue. DNA encoding a tribonection may be used in gene therapy. The present sequence represents a substantial portion of a human MSF-derived tribonection

XX Sequence 902 AA;

Query Match 48.8%; Score 3552; DB 4; Length 902;  
 Best Local Similarity 74.4%; Pred. No. 1.8e-82;  
 Matches 795; Conservative 27; Mismatches 76; Indels 170; Gaps 66;

QY 159 VKDNKKNRKTKKPTKPPVVDVDEAGSLDNGDFKVTTPDTSTTOHNKVSTSPKITTAKPIN 218  
 DB 1 VKDNKKNRKTKKPTKPPVVDVDEAGSLDNGDFKVTTPDTSTTOHNKVSTSPKITTAKPIN 60  
 QY 219 PRSLPNSDTSKETSLSLVNKKTTTETKTTTNNKQSTDGKEKTTSAKETQSIKTSK 278  
 DB 61 PRSLPNSDTSKETSLSLVNKKTTTETKTTTNNKQSTDGKEKTTSAKETQSIKTSK 120  
 QY 279 DIAPTSKVLAKPTPKAETTTKGALTTTPKEPTTTTKEPASITTPKEPTTTKSAPTPK 338  
 DB 121 DIAPTSKVLAKPTPKAETTTKGAL-----TTPKEPASITTPKEPTTTKSAPTPK 172  
 QY 339 EPAPTTKSAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 398  
 DB 173 EPAPTTKSAPTTPKEPAP--TTKEPAPTT--KEPAP--TTKEPAPTTKEPAPTT--KEPAPTT 228  
 QY 399 PKKAPPTTKEPAPTTKESPTTTPKEPAPTTTKEPAPTTTKEPAPTTAPKAPPTTKEPA 458  
 DB 229 --KEPAPTT--KEPAPTT--KEPAPTTTKEPAPTT--KEPAPTT--KEPAPTT--KEPA 281  
 QY 459 PTTKEPAPTTTKEPSPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPSPTTKE 518  
 DB 282 PTT--KEPAP--TTKEPAPTT--KEPAPTTTKEPAPTT--TTKEPAPTTTKEPAPTT--TTKE 335  
 QY 519 PAPTTPKEPAPTTPKKAPPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 578  
 DB 336 PAPTTP--KEPAPTT--KEPAPTT--KEPAPTT--TTKEPAPTT--KEPAPTT--KEPAPT 388  
 QY 579 TPCKLTPPTPEKLAPTTPKAPPTTPEELAPTTTPEEPTTPEEAPTTTPEEAPTTTPEE 638  
 DB 389 T--KEPAPTTKEP--APTKE--PAPT--KEPAPTT--KEPAPTT--KEPAPTT--KEPA 440  
 QY 639 PAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 698  
 DB 441 PAPTTP--KEPAPTT--KEPAPTT--KEPAPTT--KEPAPTT--KEPAPTT--KEPAPT 488  
 QY 699 EPTSTTSKAPPTTPKGTAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 758  
 DB 489 EPAPTKE--PAPT--KEPAPTT--KEPAPTT--KEPAPTT--KEPAPTT--KEPAPT--KEPA 541  
 QY 759 --KELAPTTTKEPSTTSKAPPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTPEPTTSV 816  
 DB 542 TTKEPAP--TTKEPAPTTKE--PAPT--KEPAPTT--KEPAPTT--KEPAPTTKE--PAPTKEP 595  
 QY 817 SPPTTTKEPTTTHKSPDSTPELSAETPKALENSP--KEPGVPTTKTPAAKPEWTTTA 874  
 DB 596 A--PTTKEPAPTTKEPAPT-----KEPAPTTTKEPAPTTKEP--APTKEPAPT----- 640

QY 875 KDKTERDLRTPPTTTAAAPMTKETATTEKTTTESKITATTQTQVISTTTQDTTTPFKITT 934  
 DB 641 -----TKEPAPTT-----KEPAPTTKEPAPTTKEPAP----- 667  
 QY 935 LKTTTLAPKVTTTKTITTTTEIMNKPBTAKPKDRAINSKATTPKQKPKAPKPKPTSTK 994  
 DB 668 -----TTKEPAPTT-----TTKEPAPTTKEP--APT--TK 684  
 QY 995 KPTMPRVRKPTTTPRKMSTMPELNPTSRIAEAMLOTTTRNQTNPNSKLVNPKSE 1054  
 DB 685 EP-----APTTPTRKMTSTMPELNPTSRIAEAML--TTTRNQTNPNSKLVNPKSE 735  
 QY 1055 DAGGAEGETHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLSDTNIENGKPVVDGLT 1114  
 DB 736 DAGGAEGETHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLSDTNIENGKPVVDGLT 795  
 QY 1115 TLRNGTLVAFRGHYFWMLSPPSPARRITEVWGIPTSDTPTTRCNCEGKTTFFFKDSQ 1174  
 DB 796 TLRNGTLVAFRGHYFWMLSPPSPARRITEVWGIPTSDTPTTRCNCEGKTTFFFKDSQ 855  
 QY 1175 YWRFTNDIKDAGYKPKPIFGFGGLTGQIVAAALSTAKYKNWPESVYFPK 1222  
 DB 856 YWRFTNDIKDAGYKPKPIFGFGGLTGQIVAAALSTA--YKNWPESVYFPK 902  
 RESULT 12  
 ABUS3253  
 ID ABUS3253 standard; protein; 551 AA.  
 XX AC ABUS3253;  
 XX AC ABUS3253;  
 DT 14-APR-2003 (first entry)  
 XX Human testes-derived DKFzphes3\_4019 homologue #2.  
 DE Human; gene therapy; vaccine; disease treatment; detection.  
 KW Homo sapiens.  
 OS Homo sapiens.  
 XX WO2001112659-A2.  
 XX 22-FEB-2001.  
 XX 18-AUG-2000; 2000WO-1B001496.  
 XX 18-AUG-1999; 99US-0149499P.  
 PR 28-SEP-1999; 99US-0156503P.  
 XX (GEHU-) GERMAN HUMAN GENOME PROJECT.  
 PA Wiemann S;  
 PI WPI; 2001-327840/34.  
 DR Nucleic acids having the sequences of clones isolated from libraries of different human tissues, useful in recombinant DNA methodologies.  
 XX Example III; Page 892-893; 1095pp; English.  
 CC This invention describes novel polynucleotides and polypeptides isolated from human cDNA libraries which can be used for gene therapy or in vaccines. The polynucleotides of the invention and antibodies encoded by them may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate polypeptide expression. The products of the invention may also be used to identify modulators of expression and activity and to down regulate expression and activity. The antibodies of the invention may also be used as diagnostic agents for detecting the presence of polypeptides in samples. This sequence represents a homologue of a polypeptide described in the disclosure of the invention  
 XX Sequence 551 AA;

Query Match 40.3%; Score 2929; DB 4; Length 551;  
Best Local Similarity 100.0%; Pred. No. 6.8e-67; Indels 0; Gaps 0;  
Matches 551; Conservative 0; Mismatches 0;

QY 556 TKKPAPTAPKEPAPTPKETAAPTPKLTPTTPEKLAAPTPBKPAAPTPPEELAPTPPEP 615  
DB 1 TKKPAPTAPKEPAPTPKETAAPTPKLTPTTPEKLAAPTPBKPAAPTPPEELAPTPPEP 60  
QY 616 TPTTPEEPAPTPKAAAPNTPKEPAPTPKPAAPTPKPAAPTPKETAAPTPKGTAPTT 675  
DB 61 TPTTPEEPAPTPKAAAPNTPKEPAPTPKPAAPTPKPAAPTPKETAAPTPKGTAPTT 120  
QY 676 LKEPAPTPKPAKAPKELAPTTTKEPTSTSDKPAAPTPKGTAPTPKPAAPTPKPAAP 735  
DB 121 LKEPAPTPKPAKAPKELAPTTTKEPTSTSDKPAAPTPKGTAPTPKPAAPTPKPAAP 180  
QY 736 TPKGTAPTTLKKEPAPTPKPAKAPKELAPTTTKEPTSTSDKPAAPTPKETAAPTPKPAAP 795  
DB 181 TPKGTAPTTLKKEPAPTPKPAKAPKELAPTTTKEPTSTSDKPAAPTPKETAAPTPKPAAP 240  
QY 796 TTPKKAPAPTPPEPPTTSVSTPTTKEPTTIHKSPDESSTPELSAEPKPALENSPKPEP 855  
DB 241 TTPKKAPAPTPPEPPTTSVSTPTTKEPTTIHKSPDESSTPELSAEPKPALENSPKPEP 300  
QY 856 GVPETTKTPAATKEPMTTAKDKTERDLRTTPETTAAAPKMTKETATTTTEKTTESKITAT 915  
DB 301 GVPETTKTPAATKEPMTTAKDKTERDLRTTPETTAAAPKMTKETATTTTEKTTESKITAT 360  
QY 916 TTQVSTSTTQDTPPFKITTLTKTTLAPKVTTTKKTTITTEIMNKPEETAKPKDRATNSKA 975  
DB 361 TTQVSTSTTQDTPPFKITTLTKTTLAPKVTTTKKTTITTEIMNKPEETAKPKDRATNSKA 420  
QY 976 TTPKPKQPTKAPKKPTSTKKPKTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTT 1035  
DB 421 TTPKPKQPTKAPKKPTSTKKPKTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTT 480  
QY 1036 TRNQTPNSKLVENPKSEDAAGAGETPHMLLRPHVFMPEVTPDMMDYLPRVFNQGLIIN 1095  
DB 481 TRNQTPNSKLVENPKSEDAAGAGETPHMLLRPHVFMPEVTPDMMDYLPRVFNQGLIIN 540  
QY 1096 PMLSDETINICN 1106  
DB 541 PMLSDETINICN 551

RESULT 13  
ABU53252  
ID ABU53252 standard; protein; 546 AA.  
XX AC ABU53252;  
XX 14-APR-2003 (first entry)  
XX Human testes-derived DKFZphes3\_4o19 homologue #1.  
XX Human; gene therapy; vaccine; disease treatment; detection.  
XX Homo sapiens.  
XX WO200112659-A2.  
XX 22-FEB-2001.  
XX 18-AUG-2000; 2000WO-IB001496.  
XX 18-AUG-1999; 99US-0149499P.  
XX 28-SEP-1999; 99US-0156503P.  
XX (GEHU-) GERMAN HUMAN GENOME PROJECT.  
XX Wiemann S;  
XX WPI; 2001-327840/34.

XX Nucleic acids having the sequences of clones isolated from libraries of  
PT different human tissues, useful in recombinant DNA methodologies.  
XX Example III; Page 892; 1095pp; English.  
XX This invention describes novel polynucleotides and polypeptides isolated  
CC from human cDNA libraries which can be used for gene therapy or in  
CC vaccines. The polynucleotides of the invention and antibodies encoded by  
CC them may be used in the prevention, diagnosis and treatment of diseases  
CC associated with inappropriate polypeptide expression. The products of the  
CC invention may also be used to identify modulators of expression and  
CC activity and to down regulate expression and activity. The antibodies of  
CC the invention may also be used as diagnostic agents for detecting the  
CC presence of polypeptides in samples. This sequence represents a homologue  
CC of a polypeptide described in the disclosure of the invention  
XX Sequence 546 AA;  
SQ

Query Match 40.1%; Score 2920; DB 4; Length 546;  
Best Local Similarity 100.0%; Pred. No. 1.1e-66;  
Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 447 KKPAPTPKPAAPTPKPAAPTPKPSPTTPKPAAPTTTKSAPTTTKKPAAPTTKSAPT 506  
DB 1 KKPAPTPKPAAPTPKPAAPTPKPSPTTPKPAAPTTTKSAPTTTKKPAAPTTKSAPT 60  
QY 507 TPKERSPTTKKPAAPTPKPAAPTPKPAAPTPKPAAPTPKPAAPTTTKKPAAPTKPE 566  
DB 61 TPKERSPTTKKPAAPTPKPAAPTPKPAAPTPKPAAPTPKPAAPTTTKKPAAPTKPE 120  
QY 567 PAPTTPKETAPTPKKLTPTTPEKLAAPTPPEKAPTTPEELAPTTPEEPPTTPEEPAPT 626  
DB 121 PAPTTPKETAPTPKKLTPTTPEKLAAPTPPEKAPTTPEELAPTTPEEPPTTPEEPAPT 180  
QY 627 TPKAAAPNTPKPAAPTPKPAAPTPKPAAPTPKPAAPTPKPAAPTTTKKPAAPTKPK 686  
DB 181 TPKAAAPNTPKPAAPTPKPAAPTPKPAAPTPKPAAPTTTKKPAAPTKPK 240  
QY 687 PAKKELAPTTTKEPTSTSDKPAAPTPKGTAPTPKPAAPTPKPAAPTTTKGTAAPTTLK 746  
DB 241 PAKKELAPTTTKEPTSTSDKPAAPTPKGTAPTPKPAAPTPKPAAPTTTKGTAAPTTLK 300  
QY 747 EPAPTPPKPAKELAPTTTKGPTSTSDKPAAPTPKETAAPTPKPAAPTPKPAAPTTTP 806  
DB 301 EPAPTPPKPAKELAPTTTKGPTSTSDKPAAPTPKETAAPTPKPAAPTPKPAAPTTTP 360  
QY 807 ETDPPTTSVSTPTTKEPTTIHKSPDESSTPELSAEPKPALENSPKPEPGVPTTKTAAAT 866  
DB 361 ETDPPTTSVSTPTTKEPTTIHKSPDESSTPELSAEPKPALENSPKPEPGVPTTKTAAAT 420  
QY 867 KPEMTTAKDKTERDLRTTPETTAAAPKMTKETATTTTEKTTESKITATTTQVSTSTTQD 926  
DB 421 KPEMTTAKDKTERDLRTTPETTAAAPKMTKETATTTTEKTTESKITATTTQVSTSTTQD 480  
QY 927 TTPFKITTLTKTTLAPKVTTTKKTTITTEIMNKPEETAKPKDRATNSKATTPKPKQPTKA 986  
DB 481 TTPFKITTLTKTTLAPKVTTTKKTTITTEIMNKPEETAKPKDRATNSKATTPKPKQPTKA 540  
QY 987 PKKPTS 992  
DB 541 PKKPTS 546

RESULT 14  
AAO18834  
ID AAO18834 standard; protein; 538 AA.  
XX AC AAO18834;  
XX 29-OCT-2002 (first entry)  
XX 3' cartilage superficial zone protein coding sequence encoded protein.

XX S2P; superficial zone protein; cartilage; lubrication; human;  
 KW degenerative joint condition; arthritis; osteoporosis; trauma; CACP;  
 KW chondroitin sulphate substitution consensus; antiarthritis;  
 KW antirheumatic; osteopathic; antigout; antiinflammatory; dermatological;  
 KW immunosuppressive.  
 XX Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 PH Misc-difference 114 /note= "encoded by ACTACT"  
 FT  
 FT  
 XX  
 XX WO200262847-A2.  
 XX 15-AUG-2002.  
 XX 31-DEC-2001; 2001WO-US050379.  
 XX 29-DEC-2000; 2000US-0258920P.  
 XX (GLAXO) GLAXO GROUP LTD.  
 XX (RUSH-) RUSH PRESBYTERIAN ST LUKE MEDICAL CENT.  
 XX Hutchins JT, Kuettner KE, Schmid TM, Schumacher BL, Su J;  
 PI Dixon EP;  
 PI WPI; 2002-636585/58.  
 DR N-PSDB; AAL49079.  
 DR  
 XX New purified superficial zone protein (SZP) polypeptides, useful for  
 PT treating degenerative joint conditions, e.g. osteoarthritis, rheumatoid  
 PT arthritis, gout, spondylarthritis, synovitis, tendonitis, lupus, or  
 PT osteoporosis.  
 XX  
 XX Claim 59; Page 86-87; 89pp; English.  
 XX  
 CC The present invention provides the protein and coding sequences of human  
 CC superficial zone protein (SZP). The protein is involved in the  
 CC lubrication of joints, and the sequences can be used in the treatment of  
 CC degenerative joint conditions or to delay symptoms of a degenerative  
 CC joint condition, e.g. osteoarthritis, rheumatoid arthritis, gout,  
 CC psoriatic arthritis, reactive arthritis, viral or post viral arthritis,  
 CC spondylarthritis, juvenile arthritis, synovitis, tendonitis, systemic  
 CC lupus erythematosus, CACP, osteoporosis or trauma. The present sequence  
 CC is the protein encoded by the human 3' cartilage SZP cDNA  
 XX  
 SQ Sequence 538 AA;  
 Query Match 39.2%; Score 2850.9; DB 5; Length 538;  
 Best Local Similarity 99.8%; Pred. No. 6.5e-65;  
 Matches 538; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
 QY 825 PTTIHKSPDESPFELSAETPKALENSPKPEGVPTTKTPAATKPEMTTAKDKTTERDLR 884  
 Db 1 PTTIHKSPDESPFELSAETPKALENSPKPEGVPTTKTPAATKPEMTTAKDKTTERDLR 60  
 QY 885 TTPETTTAAPKMTKETATTTETTESKITATTQTSTTTQDTPPKITLTKTLAPKV 944  
 Db 61 TTPETTTAAPKMTKETATTTETTESKITATTQTSTTTQDTPPKITLTK-TTLAPKV 119  
 QY 945 TTTKKTITTTETIMNKPETAKPKDRATNSKATTPPKPKPTKAPKPTSTKPKPTWPRVK 1004  
 Db 120 TTTKKTITTTETIMNKPETAKPKDRATNSKATTPPKPKPTKAPKPTSTKPKPTWPRVK 179  
 QY 1005 PKTTPPRKMTSTMPELNPTSIAEAMLTOTTRPNQTPNSKLVEVNPKSSEDAGGAEGETP 1064  
 Db 180 PKTTPPRKMTSTMPELNPTSIAEAMLTOTTRPNQTPNSKLVEVNPKSSEDAGGAEGETP 239  
 QY 1065 HMLLRPHVMPVETPDMVDYLPVPVNGIILINPMLSDETNI CNMKPVVDGLTTILRNGTLVAF 1124  
 Db 240 HMLLRPHVMPVETPDMVDYLPVPVNGIILINPMLSDETNICNGKPVVDGLTTILRNGTLVAF 299

QY 1125 RGHYFNMLSPPSPARRITEVWGIPSPIDTFTVTRCNCEGKTFFFKDSQVWRFTNDIKD 1184  
 Db 300 RGHYFNMLSPPSPARRITEVWGIPSPIDTFTVTRCNCEGKTFFFKDSQVWRFTNDIKD 359  
 QY 1185 AGYKPIFKGFGGLTGQIVAAALSTAKYKNWPESVYEFKRGSGIOQYIKQSPVOKCPGRR 1244  
 Db 360 AGYKPIFKGFGGLTGQIVAAALSTAKYKNWPESVYEFKRGSGIOQYIKQSPVOKCPGRR 419  
 QY 1245 PALNYPVYGMTQVRRRRFERAIGPSQTHIRIQYSPARLAYQDKGVLHNEVKVSIILWRG 1304  
 Db 420 PALNYPVYGMTQVRRRRFERAIGPSQTHIRIQYSPARLAYQDKGVLHNEVKVSIILWRG 479  
 QY 1305 LPNVVTSAILPNIRKPDGYDYAFSKDQYINIDVPSRTARAITTRSGQTLISKVWYNCP 1363  
 Db 480 LPNVVTSAILPNIRKPDGYDYAFSKDQYINIDVPSRTARAITTRSGQTLISKVWYNCP 538  
 RESULT 15  
 ABUS3254  
 ID ABUS3254 standard; protein; 513 AA.  
 XX  
 AC ABUS3254;  
 XX  
 DT 14-APR-2003 (first entry)  
 XX  
 DE Human testes-derived DKF3phtes3\_4019 homologue #3.  
 XX  
 KW Human; gene therapy; vaccine; disease treatment; detection.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200112659-A2.  
 DN  
 XX 22-FEB-2001.  
 PD  
 XX  
 PF 18-AUG-2000; 2000WO-IB001496.  
 XX  
 PR 18-AUG-1999; 99US-0149499P.  
 PR 28-SEP-1999; 99US-0156503P.  
 XX  
 XX (GEHU-) GERMAN HUMAN GENOME PROJECT.  
 PA  
 XX Wiemann S;  
 XX  
 DR WPI; 2001-327840/34.  
 XX  
 PT Nucleic acids having the sequences of clones isolated from libraries of  
 PT different human tissues, useful in recombinant DNA methodologies.  
 XX  
 XX Example III; Page 893; 1095pp; English.  
 XX  
 CC This invention describes novel polynucleotides and polypeptides isolated  
 CC from human cDNA libraries which can be used for gene therapy or in  
 CC vaccines. The polynucleotides of the invention and antibodies encoded by  
 CC them may be used in the prevention, diagnosis and treatment of diseases  
 CC associated with inappropriate polypeptide expression. The products of the  
 CC invention may also be used to identify modulators of expression and  
 CC activity and to down regulate expression and activity. The antibodies of  
 CC the invention may also be used as diagnostic agents for detecting the  
 CC presence of polypeptides in samples. This sequence represents a homologue  
 CC of a polypeptide described in the disclosure of the invention  
 XX  
 SQ Sequence 513 AA;  
 Query Match 37.9%; Score 2757; DB 4; Length 513;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-62;  
 Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 166 RTKKKPTKPPVDEAGSLDNGDFKVTTPDTSTTQHNKVSPTKITTAKPINRPSLPP 225  
 Db 1 RTKKKPTKPPVDEAGSLDNGDFKVTTPDTSTTQHNKVSPTKITTAKPINRPSLPP 60  
 QY 226 NSDTSKETSILVNKETTETTTTNTKQTDGKEKTSKETSQTSIEKTSKADLAPTSK 285

Db	61	NSDTSKETS	LTWNKET	TVETKETT	NTNKO	TSIDG	KEKTS	SAKET	QSI	EKISAK	DLAPT	SK	120	
Qy	286	VLA	KPTPKA	ETTT	KG	PALT	TPKE	PTPT	TPKE	PA	STTP	KEPT	PTTT	345
Db	121	VLA	KPTPKA	ETTT	KG	PALT	TPKE	PTPT	TPKE	PA	STTP	KEPT	PTTT	180
Qy	346	KS	APT	TPKE	PA	STTP	TPKE	PA	STTP	TPKE	PA	STTP	TPKE	405
Db	181	KS	APT	TPKE	PA	STTP	TPKE	PA	STTP	TPKE	PA	STTP	TPKE	240
Qy	406	TP	KE	PA	STTP	TPKE	PA	STTP	TPKE	PA	STTP	TPKE	PA	465
Db	241	TP	KE	PA	STTP	TPKE	PA	STTP	TPKE	PA	STTP	TPKE	PA	300
Qy	466	AP	TT	TPKE	PA	STTP	TPKE	PA	STTP	TPKE	PA	STTP	TPKE	525
Db	301	AP	TT	TPKE	PA	STTP	TPKE	PA	STTP	TPKE	PA	STTP	TPKE	360
Qy	526	EP	APT	TPKE	PA	STTP	TPKE	PA	STTP	TPKE	PA	STTP	TPKE	585
Db	361	EP	APT	TPKE	PA	STTP	TPKE	PA	STTP	TPKE	PA	STTP	TPKE	420
Qy	586	TP	KE	PA	STTP	TPKE	PA	STTP	TPKE	PA	STTP	TPKE	PA	645
Db	421	TP	KE	PA	STTP	TPKE	PA	STTP	TPKE	PA	STTP	TPKE	PA	480
Qy	646	EP	APT	TPKE	PA	STTP	TPKE	PA	STTP	TPKE	PA	STTP	TPKE	678
Db	481	EP	APT	TPKE	PA	STTP	TPKE	PA	STTP	TPKE	PA	STTP	TPKE	513

Search completed: October 13, 2004, 11:37:15  
Job time : 116.889 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 23.693 Seconds  
(without alignments)  
3815.116 Million cell updates/sec

Title: SEQ1-F  
Perfect score: 7276

Sequence: 1 MAWKTLPIYLLLLSVFVIQ.....ARAITRSGTSLKWNVNC 1363

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.1

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/ptodata/1/iaa/5A COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PTCUS COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7276	100.0	1363	4	US-07-757-022B-52
2	7261.9	99.8	1404	4	US-07-757-022B-2
3	7261.9	99.8	1404	4	US-07-757-022B-62
4	7261.9	99.8	1404	4	US-09-298-970A-1
5	7232.9	99.4	1404	4	US-10-164-595-78
6	7058.7	97.0	1320	4	US-07-757-022B-46
7	7058.7	97.0	1320	4	US-07-757-022B-60
8	7044.6	96.8	1361	4	US-07-757-022B-40
9	7043.7	96.8	1320	4	US-10-164-595-58
10	7002	96.2	1313	4	US-07-757-022B-142
11	6987.9	96.0	1354	4	US-07-757-022B-48
12	6878.9	94.5	1314	4	US-07-757-022B-50
13	6794.7	93.4	1270	4	US-07-757-022B-44
14	6780.6	93.2	1311	4	US-07-757-022B-42
15	5820.9	80.0	1140	4	US-07-757-022B-104
16	5561	76.4	1049	4	US-07-757-022B-58
17	5495.5	75.5	1038	4	US-07-757-022B-74
18	5231.9	71.9	1022	4	US-07-757-022B-84
19	5011	68.9	941	4	US-07-757-022B-34
20	2160.9	29.7	422	4	US-07-757-022B-68
21	2146.8	29.5	463	4	US-07-757-022B-54
22	2032.8	27.9	423	4	US-07-757-022B-66
23	1886.9	25.9	372	4	US-07-757-022B-64
24	1371	18.8	5179	4	US-09-538-092-1258
25	1200.1	16.5	296	4	US-07-757-022B-70
26	1196.3	16.4	8991	4	US-08-714-741-32
27	946.2	13.0	237	4	US-07-757-022B-72

28 923.9 12.7 220 4 US-07-757-022B-96  
29 895.7 12.3 2972 3 US-09-579-181-2  
30 895.7 12.3 3118 3 US-09-579-181-1  
31 876.3 12.0 3256 4 US-09-919-172-98  
32 876.3 12.0 3256 4 US-09-976-594-22  
33 876.3 12.0 3256 4 US-09-919-039-21  
34 874.2 12.0 217 4 US-07-757-022B-76  
35 864.6 11.9 207 4 US-07-757-022B-116  
36 864.6 11.9 207 4 US-07-757-022B-136  
37 861.9 11.8 209 4 US-07-757-022B-94  
38 857.6 11.8 231 4 US-07-757-022B-30  
39 856.9 11.8 208 4 US-07-757-022B-132  
40 849 11.7 2142 4 US-09-538-092-1142  
41 835.9 11.5 204 4 US-07-757-022B-92  
42 835 11.5 1837 3 US-08-928-361B-5  
43 835 11.5 1837 4 US-09-588-995A-5  
44 829.7 11.4 4019 4 US-09-854-133-425  
45 792.8 10.9 1721 3 US-08-700-651-5

## ALIGNMENTS

### RESULT 1

US-07-757-022B-52  
; Sequence 52, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 52:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1363 amino acids  
; TYPE: AMINO ACID

Sequence 96, Appl  
Sequence 2, Appl  
Sequence 1, Appl  
Sequence 98, Appl  
Sequence 22, Appl  
Sequence 21, Appl  
Sequence 76, Appl  
Sequence 116, Appl  
Sequence 136, Appl  
Sequence 94, Appl  
Sequence 30, Appl  
Sequence 132, Appl  
Sequence 1142, Appl  
Sequence 92, Appl  
Sequence 5, Appl  
Sequence 5, Appl  
Sequence 425, Appl  
Sequence 5, Appl



QY	860	TKTAAATKPEMTTITAKDKTTTERDLRTTPETTTAAPKXTKETATTTEKITESKITATTTOV	919
Dd			
Dd	901	TKTAAATKPEMTTITAKDKTTTERDLRTTPETTTAAPKXTKETATTTEKITESKITATTTOV	960
QY	920	TSITTQDITTFPKIITLTKITTLAPKVITTTKKIIITTIEMNKPEETAAPKDRATNSKATTPK	979
Dd			
Dd	961	TSITTQDITTFPKIITLTKITTLAPKVITTTKKIIITTIEMNKPEETAAPKDRATNSKATTPK	1020
QY	980	POKETKAPKEPTSTKKPKTMPRVRKPKTTTPRKMSTMPELNPTSRIAEAMLQTTRPN	1039
Dd			
Dd	1021	POKETKAPKEPTSTKKPKTMPRVRKPKTTTPRKMSTMPELNPTSRIAEAMLQTTRPN	1080
QY	1040	QTPNSKLVEVNPXSSEDAGGAEGETBHWLLRPHVFMPEVTDPMDYLPRVPNOGIIINPMLS	1099
Dd			
Dd	1081	QTPNSKLVEVNPXSSEDAGGAEGETBHWLLRPHVFMPEVTDPMDYLPRVPNOGIIINPMLS	1140
QY	1100	DETNICNGKPVDDGLITLRNGTLVAFRGHYFWMLSFPSPPSPARRITEVWGIPSDIVFT	1159
Dd			
Dd	1141	DETNICNGKPVDDGLITLRNGTLVAFRGHYFWMLSFPSPPSPARRITEVWGIPSDIVFT	1200
QY	1160	RCNCEGTFFFKDSQWRFTNDIKDGYKPPIKFGFGLTGQIVAALSTAKYKNWPESVY	1219
Dd			
Dd	1201	RCNCEGTFFFKDSQWRFTNDIKDGYKPPIKFGFGLTGQIVAALSTAKYKNWPESVY	1260
QY	1220	FFRGSGSIQQIYYKQPVOVKCGRRPALANPVYGEMTOVRRRRFERAIGPSQTHIRIQY	1279
Dd			
Dd	1261	FFRGSGSIQQIYYKQPVOVKCGRRPALANPVYGEMTOVRRRRFERAIGPSQTHIRIQY	1320
QY	1280	SPARLAYQDKGVILHNEVKVSILLWRLGNVVTSAISLPNIKRKPDGYDYVAFSKQYYNIDV	1339
Dd			
Dd	1321	SPARLAYQDKGVILHNEVKVSILLWRLGNVVTSAISLPNIKRKPDGYDYVAFSKQYYNIDV	1380
QY	1340	PSRTARAITSRGOTLSKWYNCP	1363
Dd			
Dd	1381	PSRTARAITSRGOTLSKWYNCP	1404

RESULT 3  
US-07-757-022B-62  
; Sequence 62, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989

PRIOR APPLICATION DATA:		
; APPLICATION NUMBER: US 07/390,901		
; FILING DATE: 08-AUG-1989		
; ATTORNEY/AGENT INFORMATION:		
; NAME: Cseri, Luann		
; REGISTRATION NUMBER: 31,822		
; REFERENCE/DOCKET NUMBER: GI 5190		
; TELECOMMUNICATION INFORMATION:		
; TELEPHONE: (617)876-1170		
; TELEFAX: (617)876-5851		
; INFORMATION FOR SEQ ID NO: 62:		
; SEQUENCE CHARACTERISTICS:		
; LENGTH: 1404 amino acids		
; TYPE: AMINO ACID		
; TOPOLOGY: linear		
; MOLECULE TYPE: protein		
US-07-757-022B-62		
Query Match 99.8%; Score 7261.9; DB 4; Length 1404;		
Best Local Similarity 97.1%; Pred. No. 4.2e-196;		
Matches 1363; Conservative 0; Mismatches 0; Indels 41; Gaps 1;		
QY	1	MAWKTLPIYLILLLSVFVIQVSSQ-----25
DB	1	MAWKTLPIYLILLLSVFVIQVSSQDLSSCAGRCGEGYSDATCNC DYNCQHYMECCPDF 60
QY	26	-----ELSCCKGRCFESFERGREGDCDCAQCKKYDKCCPDYESFCAEVHNPTSPSSKKAP 79
DB	61	KRVCTAELSCGRCFESFERGREGDCDCAQCKKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
QY	80	PPSGASQTKSTTKRSPKPNKKTKVIESEBITEHSHSVSENOESSSSSSSSSTIW 139
DB	121	PPSGASQTKISTTKRSPKPNKKTKVIESEBITEHSHSVSENOESSSSSSSSSTIW 180
QY	140	KIKSSKNSAANRELQKKLVKDNKNRKKKPKPPVVDVDEAGSLDNGDFKVTTPDTST 199
DB	181	KIKSSKNSAANRELQKKLVKDNKNRKKKPKPPVVDVDEAGSLDNGDFKVTTPDTST 240
QY	200	TQHNKSVTSPIITAKPINRPSLPNPSDTSKETSLSLVNKETTVETKETTINKQTSIDG 259
DB	241	TQHNKSVTSPIITAKPINRPSLPNPSDTSKETSLSLVNKETTVETKETTINKQTSIDG 300
QY	260	KEKTSKETSQTSKETSADLAPTSKVLAKPTKAEITTKGPALTTPEKPTPTPKPEPAS 319
DB	301	KEKTSKETSQTSKETSADLAPTSKVLAKPTKAEITTKGPALTTPEKPTPTPKPEPAS 360
QY	320	TTPEKPTPTTKSAPTTPKPEAPTTTKSAPTTPKPEAPTTTKPEAPTTTKPEAPTTTKPE 379
DB	361	TTPEKPTPTTKSAPTTPKPEAPTTTKSAPTTPKPEAPTTTKPEAPTTTKPEAPTTTKPE 420
QY	380	APTPTKSAPTTPKPEAPTTPKKAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPK 439
DB	421	APTPTKSAPTTPKPEAPTTPKKAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPK 480
QY	440	EPAPTAKKAPTTPKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTTKPEAPT 499
DB	481	EPAPTAKKAPTTPKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTTKPEAPT 540
QY	500	TTKSAPTTPKPEPSTTTKEAPTTPKPEAPTTPKKAPTTPKPEAPTTPKPEAPTTTKKP 559
DB	541	TTKSAPTTPKPEPSTTTKEAPTTPKPEAPTTPKKAPTTPKPEAPTTPKPEAPTTTKKP 600
QY	560	APTAPKPEAPTTPKETAPTTPKLTTPPEKLAETTPPEKAPTTPEELAPTTPEEPTPT 619
DB	601	APTAPKPEAPTTPKETAPTTPKLTTPPEKLAETTPPEKAPTTPEELAPTTPEEPTPT 660
QY	620	PEEPAPTTPKAAADNTPKPEAPTTPKPEAPTTPKETAPTTPKGTAPTTLKPE 679
DB	661	PEEPAPTTPKAAADNTPKPEAPTTPKPEAPTTPKETAPTTPKGTAPTTLKPE 720
QY	680	APTTPKKPAPKELAPTTTKETSTSDKPAETTPKGTAPTTPKPEAPTTPKPEAPTTPKG 739
DB	721	APTTPKKPAPKELAPTTTKETSTSDKPAETTPKGTAPTTPKPEAPTTPKPEAPTTPKG 780
QY	740	TAPTTLKBPAPTTPKKAPKELAPTTTKGPTSTTSDKPAETTPKETAPTTPKPEAPTTPK 799
DB	781	TAPTTLKBPAPTTPKKAPKELAPTTTKGPTSTTSDKPAETTPKETAPTTPKPEAPTTPK 840
QY	800	KDAPTTPETPTTSEVSTPTTTKEPTTIHKSPDSESTPSELSAETPPKALENSPKKEGVT 859
DB	841	KDAPTTPETPTTSEVSTPTTTKEPTTIHKSPDSESTPSELSAETPPKALENSPKKEGVT 900
QY	860	TKTPAATKPEMTTAKDKTTERDLBETTPETTTAAKPKMTKETATTTTEKTTESKITATTQV 919
DB	901	TKTPAATKPEMTTAKDKTTERDLBETTPETTTAAKPKMTKETATTTTEKTTESKITATTQV 960
QY	920	TSTTTQDTPPKITTLKTTLAPKVTITTKTITTEIMNKPBEETAKPKDRATNSKATTPK 979
DB	961	TSTTTQDTPPKITTLKTTLAPKVTITTKTITTEIMNKPBEETAKPKDRATNSKATTPK 1020
QY	980	POKPTKAPKKPTSTKPKTMPVRKPTTPTRKMTSTMPELNPTSPRIAEAMLQTTTRPN 1039
DB	1021	POKPTKAPKKPTSTKPKTMPVRKPTTPTRKMTSTMPELNPTSPRIAEAMLQTTTRPN 1080
QY	1040	QTPNSKLVEVNPKSEDAGAGESTPHMLLRPHVFMPEVTPDMDYLRVNPQGIINPMLS 1099
DB	1081	QTPNSKLVEVNPKSEDAGAGESTPHMLLRPHVFMPEVTPDMDYLRVNPQGIINPMLS 1140
QY	1100	DETNICNGKPVDCGLTTLRNGTLVAFRGHYFWMLSPPSPSPARRITEVMGIPSPIDTVFT 1159
DB	1141	DETNICNGKPVDCGLTTLRNGTLVAFRGHYFWMLSPPSPSPARRITEVMGIPSPIDTVFT 1200
QY	1160	RCNCEGKTFPFDQSWRFTNDIKDAGYKPKPIFKGFGGLTGQIVAAALSTAKYKNWPESVY 1219
DB	1201	RCNCEGKTFPFDQSWRFTNDIKDAGYKPKPIFKGFGGLTGQIVAAALSTAKYKNWPESVY 1260
QY	1220	FFKRGGSIOQYIVKQBPVQKCPGRRPALNVPVVGEMTQVRRRRFERAIGPSQHTTIRIQY 1279
DB	1261	FFKRGGSIOQYIVKQBPVQKCPGRRPALNVPVVGEMTQVRRRRFERAIGPSQHTTIRIQY 1320
QY	1280	SPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLPNIRKPDGYDYIYAFSKDQYNNIDV 1339
DB	1321	SPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLPNIRKPDGYDYIYAFSKDQYNNIDV 1380
QY	1340	PSRTARAITRSGQTLSKVWYNCP 1363
DB	1381	PSRTARAITRSGQTLSKVWYNCP 1404
RESULT 4		
US-09-298-970A-1		
; Sequence 1, Application US/09298970A		
; Patent No. 6743774		
; GENERAL INFORMATION:		
; APPLICANT: Jay, Gregory D.		
; TITLE OF INVENTION: TRIBONECTINS		
; FILE REFERENCE: 21486-026		
; CURRENT APPLICATION NUMBER: US/09/298,970A		
; CURRENT FILING DATE: 2001-06-19		
; NUMBER OF SEQ ID NOS: 16		
; SOFTWARE: FastSeq for Windows Version 4.0		
; SEQ ID NO 1		
; LENGTH: 1404		
; TYPE: PRT		
; ORGANISM: Homo sapiens		
US-09-298-970A-1		
Query Match 99.8%; Score 7261.9; DB 4; Length 1404;		
Best Local Similarity 97.1%; Pred. No. 4.2e-196;		
Matches 1363; Conservative 0; Mismatches 0; Indels 41; Gaps 1;		
QY	1	MAWKTLPIYLILLLSVFVIQVSSQ-----25
DB	1	MAWKTLPIYLILLLSVFVIQVSSQDLSSCAGRCGEGYSDATCNC DYNCQHYMECCPDF 60
QY	26	-----ELSCCKGRCFESFERGREGDCDCAQCKKYDKCCPDYESFCAEVHNPTSPSSKKAP 79
DB	61	KRVCTAELSCGRCFESFERGREGDCDCAQCKKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
QY	80	PPSGASQTKSTTKRSPKPNKKTKVIESEBITEHSHSVSENOESSSSSSSSSTIW 139
DB	121	PPSGASQTKISTTKRSPKPNKKTKVIESEBITEHSHSVSENOESSSSSSSSSTIW 180
QY	140	KIKSSKNSAANRELQKKLVKDNKNRKKKPKPPVVDVDEAGSLDNGDFKVTTPDTST 199
DB	181	KIKSSKNSAANRELQKKLVKDNKNRKKKPKPPVVDVDEAGSLDNGDFKVTTPDTST 240
QY	200	TQHNKSVTSPIITAKPINRPSLPNPSDTSKETSLSLVNKETTVETKETTINKQTSIDG 259
DB	241	TQHNKSVTSPIITAKPINRPSLPNPSDTSKETSLSLVNKETTVETKETTINKQTSIDG 300
QY	260	KEKTSKETSQTSKETSADLAPTSKVLAKPTKAEITTKGPALTTPEKPTPTPKPEPAS 319
DB	301	KEKTSKETSQTSKETSADLAPTSKVLAKPTKAEITTKGPALTTPEKPTPTPKPEPAS 360
QY	320	TTPEKPTPTTKSAPTTPKPEAPTTTKSAPTTPKPEAPTTTKPEAPTTTKPEAPTTTKPE 379
DB	361	TTPEKPTPTTKSAPTTPKPEAPTTTKSAPTTPKPEAPTTTKPEAPTTTKPEAPTTTKPE 420
QY	380	APTPTKSAPTTPKPEAPTTPKKAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPK 439
DB	421	APTPTKSAPTTPKPEAPTTPKKAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPK 480
QY	440	EPAPTAKKAPTTPKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTTKPEAPT 499
DB	481	EPAPTAKKAPTTPKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTTKPEAPT 540
QY	500	TTKSAPTTPKPEPSTTTKEAPTTPKPEAPTTPKKAPTTPKPEAPTTPKPEAPTTTKKP 559
DB	541	TTKSAPTTPKPEPSTTTKEAPTTPKPEAPTTPKKAPTTPKPEAPTTPKPEAPTTTKKP 600
QY	560	APTAPKPEAPTTPKETAPTTPKLTTPPEKLAETTPPEKAPTTPEELAPTTPEEPTPT 619
DB	601	APTAPKPEAPTTPKETAPTTPKLTTPPEKLAETTPPEKAPTTPEELAPTTPEEPTPT 660
QY	620	PEEPAPTTPKAAADNTPKPEAPTTPKPEAPTTPKETAPTTPKGTAPTTLKPE 679
DB	661	PEEPAPTTPKAAADNTPKPEAPTTPKPEAPTTPKETAPTTPKGTAPTTLKPE 720
QY	680	APTTPKKPAPKELAPTTTKETSTSDKPAETTPKGTAPTTPKPEAPTTPKPEAPTTPKG 739
DB	721	APTTPKKPAPKELAPTTTKETSTSDKPAETTPKGTAPTTPKPEAPTTPKPEAPTTPKG 780



Db 61 KRVCTAELSCGRCFESFERGECDDAQQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120  
Qy 80 PPSGASQTIKSTTKRSKPPNNKTKKVIIESEITEEHSVSENQESSSSSSSTIWM 139  
Db 121 PPSGASQTIKSTTKRSKPPNNKTKKVIIESEITEEHSVSENQESSSSSSSTIWM 180  
Qy 140 KIKSSKNSAANRELQKKLVKDNKNRTKKKPTPKPPVVDGAGSLDNGDFKVTPTDST 199  
Db 181 KIKSSKNSAANRELQKKLVKDNKNRTKKKPTPKPPVVDGAGSLDNGDFKVTPTDST 240  
Qy 200 TQHNKVSSTPKITTTAKINPRPSLPNSDDTSKETSITVKNKETTVEKTTNNKQTSIDG 259  
Db 241 TQHNKVSSTPKITTTAKINPRPSLPNSDDTSKETSITVKNKETTVEKTTNNKQTSIDG 300  
Qy 260 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTKEPTTTPKEPAS 319  
Db 301 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTKEPTTTPKEPAS 360  
Qy 320 TTPKEPTTTIKSAPTTTKEPATTTTKSAPTTTKEPATTTTKEPATTTTKEPATTTTKEP 379  
Db 361 TTPKEPTTTIKSAPTTTKEPATTTTKSAPTTTKEPATTTTKEPATTTTKEPATTTTKEP 420  
Qy 380 APATTKSAPTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTPK 439  
Db 421 APATTKSAPTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTPK 480  
Qy 440 EPAPTAPKKAPATTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPAT 499  
Db 481 EPAPTAPKKAPATTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPAT 540  
Qy 500 TTKSAPTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEP 559  
Db 541 TTKSAPTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEP 600  
Qy 560 APATAPKEPATTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTT 619  
Db 601 APATAPKEPATTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTT 660  
Qy 620 PSEPAPTTTKAAPNTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPAT 679  
Db 661 PSEPAPTTTKAAPNTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPAT 720  
Qy 680 APATPKAPKELAPTTTKEPTTSTSDKAPATTTPKGTAPTTTKEPATTTTKEPATTTTKEP 739  
Db 721 APATPKAPKELAPTTTKEPTTSTSDKAPATTTPKGTAPTTTKEPATTTTKEPATTTTKEP 780  
Qy 740 TAPTTTKEPATTTPKKAPKELAPTTTKEPTTSTSDKAPATTTPKGTAPTTTKEPATTTTKEP 799  
Db 781 TAPTTTKEPATTTPKKAPKELAPTTTKEPTTSTSDKAPATTTPKGTAPTTTKEPATTTTKEP 840  
Qy 800 KPAPTTTPEPTTSEVSTPTTKEPTTTHKSPDESTPELSAEPPTKALENSGKPCVPT 859  
Db 841 KPAPTTTPEPTTSEVSTPTTKEPTTTHKSPDESTPELSAEPPTKALENSGKPCVPT 900  
Qy 860 TKTPAATKPEMTTAKOKTTERDLRTTPTTTAAKPMKETAATTTKEKTTESKITATTQV 919  
Db 901 TKTPAATKPEMTTAKOKTTERDLRTTPTTTAAKPMKETAATTTKEKTTESKITATTQV 960  
Qy 920 TSTTTQDTPPKITLTKTTLAPKVTTHKGTITTTTEIMNKPEETAKPKDRATNSKATTPK 979  
Db 961 TSTTTQDTPPKITLTKTTLAPKVTTHKGTITTTTEIMNKPEETAKPKDRATNSKATTPK 1020  
Qy 980 POKPTKAPKPTSTTKPKTMPVRPKTPTTPRKMTSTNPELNPTSRIAEAMLOTTTRPN 1039  
Db 1021 POKPTKAPKPTSTTKPKTMPVRPKTPTTPRKMTSTNPELNPTSRIAEAMLOTTTRPN 1080  
Qy 1040 QTPNSKLVENPKSEDAGAGETPHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS 1099  
Db 1081 QTPNSKLVENPKSEDAGAGETPHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS 1140  
Qy 1100 DETNICNGKPVUGLTTLRNGTLVAFRGHYFWMLSFPSPARRITEVWGIPSPIDTFT 1159

Db 1141 DETNICNGKPVUGLTTLRNGTLVAFRGHYFWMLSFPSPARRITEVWGIPSPIDTFT 1200  
Qy 1160 RCNCEGKTFPFKDSQYWRFTNDIKDAGYPKPIFKGFGGLTGQIYAALSTAKYKNWPESVY 1219  
Db 1201 RCNCEGKTFPFKDSQYWRFTNDIKDAGYPKPIFKGFGGLTGQIYAALSTAKYKNWPESVY 1260  
Qy 1220 PFKEGGSIOQYIYKQEPVQKCPGRPALNTPVYVYVYVYVYVYVYVYVYVYVYVYVY 1279  
Db 1261 PFKEGGSIOQYIYKQEPVQKCPGRPALNTPVYVYVYVYVYVYVYVYVYVYVYVY 1320  
Qy 1280 SPARLAYQDKGVLHNEVKVSLMRGLNVTSAISLNPRIKPDGQYDYAFSKDQYYNIDV 1339  
Db 1321 SPARLAYQDKGVLHNEVKVSLMRGLNVTSAISLNPRIKPDGQYDYAFSKDQYYNIDV 1380  
Qy 1340 PSRTARAITTRSGQTLSSKVMYNCP 1363  
Db 1381 PSRTARAITTRSGQTLSSKVMYNCP 1404

RESULT 5  
US-10-164-595-78  
; Sequence 78, Application US/10164595  
; Patent No. 6657054  
; GENERAL INFORMATION:  
; APPLICANT: Origene Technologies, Inc  
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides  
; FILE REFERENCE: IU 103 R1  
; CURRENT APPLICATION NUMBER: US/10/164,595  
; CURRENT FILING DATE: 2002-06-10  
; NUMBER OF SEQ ID NOS: 80  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO: 78  
; LENGTH: 1404  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-164-595-78

Query Match 99.4%; Score 7232.9; DB 4; Length 1404;  
Best Local Similarity 96.8%; Pred. No. 2.7e-195;  
Matches 1359; Conservative 0; Mismatches 4; Indels 41; Gaps 1;

Qy 1 MAWKTLPIYLLLLSVFVIQQVSSQ----- 25  
Db 1 MAWKTLPIYLLLLSVFVIQQVSSQDLSSCAGRCGEGYGRDATCNCDYNCQHYMECCPDF 60  
Qy 26 -----ELSCGRCFESFERGECDDAQQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 79  
Db 61 KRVCTAELSCGRCFESFERGECDDAQQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120  
Qy 80 PPSGASQTIKSTTKRSKPPNNKTKKVIIESEITEEHSVSENQESSSSSSSSSTIWM 139  
Db 121 PPSGASQTIKSTTKRSKPPNNKTKKVIIESEITEEHSVSENQESSSSSSSSSTIWM 180  
Qy 140 KIKSSKNSAANRELQKKLVKDNKNRTKKKPTPKPPVVDGAGSLDNGDFKVTPTDST 199  
Db 181 KIKSSKNSAANRELQKKLVKDNKNRTKKKPTPKPPVVDGAGSLDNGDFKVTPTDST 240  
Qy 200 TQHNKVSSTPKITTTAKINPRPSLPNSDDTSKETSITVKNKETTVEKTTNNKQTSIDG 259  
Db 241 TQHNKVSSTPKITTTAKINPRPSLPNSDDTSKETSITVKNKETTVEKTTNNKQTSIDG 300  
Qy 260 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTKEPTTTPKEPAS 319  
Db 301 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTKEPTTTPKEPAS 360  
Qy 320 TTPKEPTTTIKSAPTTTKEPATTTTKSAPTTTKEPATTTTKEPATTTTKEPATTTTKEP 379  
Db 361 TTPKEPTTTIKSAPTTTKEPATTTTKSAPTTTKEPATTTTKEPATTTTKEPATTTTKEP 420  
Qy 380 APATTKSAPTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTPK 439  
Db 421 APATTKSAPTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTPK 480



```
QY 361 KEPAATTTKEPAATTTKEPAATTTKSAATTTKEPAATTTKEPAATTTKEPAATTTKEPTP 420
Db 318 KEPAATTTKEPAATTTKEPAATTTKSAATTTKEPAATTTKEPAATTTKEPAATTTKEPTP 377
QY 421 TTKEPAATTTKEPAATTTKEPAATTTKAPKAPATTTKEPAATTTKEPAATTTKEPSPTTKE 480
Db 378 TTKEPAATTTKEPAATTTKEPAATTTKAPKAPATTTKEPAATTTKEPAATTTKEPSPTTKE 437
QY 481 PAATTTKSAATTTKEPAATTTKSAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTT 540
Db 438 PAATTTKSAATTTKEPAATTTKSAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTT 497
QY 541 KEPAATTTKEPAATTTKAPATTAKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPA 600
Db 498 KEPAATTTKEPAATTTKAPATTAKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPA 557
QY 601 PTTPEELAPTTPEPTPTTPEEPATTTKAAAPNTKEPAATTTKEPAATTTKEPAATTTKEPA 660
Db 558 PTTPEELAPTTPEPTPTTPEEPATTTKAAAPNTKEPAATTTKEPAATTTKEPAATTTKEPA 617
QY 661 KETAPTTKGTAPTTKEPAATTTKAPKAPKELAPTTKEPTSTTSDKPAATTTKGTAPT 720
Db 618 KETAPTTKGTAPTTKEPAATTTKAPKAPKELAPTTKEPTSTTSDKPAATTTKGTAPT 677
QY 721 KEPAATTTKEPAATTTKGTAPTTKEPAATTTKAPKAPKELAPTTKGTSTTSDKPAAT 780
Db 678 KEPAATTTKEPAATTTKGTAPTTKEPAATTTKAPKAPKELAPTTKGTSTTSDKPAAT 737
QY 781 TPKEATTTKEPAATTTKEPAATTTPEPTPTTSEVSTPTTKEPTTIHKSPDESTPELS 840
Db 738 TPKEATTTKEPAATTTKEPAATTTPEPTPTTSEVSTPTTKEPTTIHKSPDESTPELS 797
QY 841 AEPTPKALENSKEPGVPTTKTAAATPEMTTAKDKTTERDLRTTPEPTTAAAPKMTKET 900
Db 798 AEPTPKALENSKEPGVPTTKTAAATPEMTTAKDKTTERDLRTTPEPTTAAAPKMTKET 857
QY 901 AITTEKTTESKITATTQVTTSTTQDTTPPKITTLKTITTLAPKVTTTKITITTEIMNKP 960
Db 858 AITTEKTTESKITATTQVTTSTTQDTTPPKITTLKTITTLAPKVTTTKITITTEIMNKP 917
QY 961 EETAKPKDRATNSKATTPKPKETKAPKPTSTKPKTMRPRVKPXTTTPRKTSTMP 1020
Db 918 EETAKPKDRATNSKATTPKPKETKAPKPTSTKPKTMRPRVKPXTTTPRKTSTMP 977
QY 1021 LNPTSRIAEAMLOTTTRPQTPNSKLVEVNPXSEDAGGAEGETPHMLLRPHVPMPEVTPD 1080
Db 978 LNPTSRIAEAMLOTTTRPQTPNSKLVEVNPXSEDAGGAEGETPHMLLRPHVPMPEVTPD 1037
QY 1081 MDYLRVPVNOGIIINPMLSDETNI CNKPKVDGLTTLRNGTLVAFRGHYFWMLSPFSPSP 1140
Db 1038 MDYLRVPVNOGIIINPMLSDETNI CNKPKVDGLTTLRNGTLVAFRGHYFWMLSPFSPSP 1097
QY 1141 ARRIITEVWGIPSPIDTVTRCNCBGTFFFKDSQYWRFTNDIKDAGYKPIPKFGGGLTG 1200
Db 1098 ARRIITEVWGIPSPIDTVTRCNCBGTFFFKDSQYWRFTNDIKDAGYKPIPKFGGGLTG 1157
QY 1201 QIVAAALSTAKYKNWPESVYFFKRGGSIQYIYKQEPVQKCGRRPALNVPVGMTOVER 1260
Db 1158 QIVAAALSTAKYKNWPESVYFFKRGGSIQYIYKQEPVQKCGRRPALNVPVGMTOVER 1217
QY 1261 RFPERAIGPSQTHTRI QVSPARLAYQDKGVHLNEVKVSI LWRGLPNVVTSAISLPNIRK 1320
Db 1218 RFPERAIGPSQTHTRI QVSPARLAYQDKGVHLNEVKVSI LWRGLPNVVTSAISLPNIRK 1277
QY 1321 PDGYDYAFSKDQYNYNIDVPSRTARAITTRSQTL SKVWYNCP 1363
Db 1278 PDGYDYAFSKDQYNYNIDVPSRTARAITTRSQTL SKVWYNCP 1320
```

RESULT 7

US-07-757-022B-60

; Sequence 60, Application US/07757022B

```
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseerr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1320 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-60
```

```
Query Match 97.0%; Score 7058.7; DB 4; Length 1320;
Best Local Similarity 96.8%; Pred. No. 1.9e-190;
Matches 1320; Conservative 0; Mismatches 0; Indels 43; Gaps 1;
```

```
QY 1 MAWKTLPIYLLLLLSVFVIQQVSSOELSCGRCFSPFERGRCDCDAQCKKYDKCCPDYE 60
Db 1 MAWKTLPIYLLLLLSVFVIQQVSSOELSCGRCFSPFERGRCDCDAQCKKYDKCCPDYE 60
QY 61 SFCAEVHNPTSPSSKAPPPSGASQTIKSTTKSPKPNKKTKKVIKESBEITEHSVS 120
Db 61 SFCAEVHNPTSPSSKAPPPSGASQTIKSTTKSPKPNKKTKKVIKESBEITEHSVS 115
QY 121 ENQESSSSSSSSSSSTIWKIKSSKNKSAANRELQKLVKDKNKKRTKKKTKPKPPVDE 180
Db 116 -----VKDNKKRTKKKTKPKPPVDE 137
QY 181 AGSGLNDGDFKVTTPDTSTTQHNKVSTSPKITTAKPINPRLSPNSDTSKETSLTVNKE 240
Db 138 AGSGLNDGDFKVTTPDTSTTQHNKVSTSPKITTAKPINPRLSPNSDTSKETSLTVNKE 197
QY 241 TTVTETKTTTNTKQSTTDGKEKTTSAKETQSIKTSKADLAPTSKVLAKPTEKATTTKG 300
```

Db 198 TTVEKETTNNKQTSQCKEKTTSKAKETQSIKTSKDLAPTQVLAKPTPKASTTNG 257

QY 301 PALTTPKEPTPTTPKEPASTTTPKEPTPTTIKSAPTTPKEPAPTTPKSAPTTPKEPAPTTP 360

Db 258 PALTTPKEPTPTTPKEPASTTTPKEPTPTTIKSAPTTPKEPAPTTPKSAPTTPKEPAPTTP 317

QY 361 KEPAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKEPTTP 420

Db 318 KEPAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKEPTTP 377

QY 421 TTPKEPAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKEPTTP 480

Db 378 TTPKEPAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKEPTTP 437

QY 481 PAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKEPTTP 540

Db 438 PAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKEPTTP 497

QY 541 KGPAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKEPTTP 600

Db 498 KGPAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKEPTTP 557

QY 601 PTTPEELAPTTPKEPTPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKEPTTP 660

Db 558 PTTPEELAPTTPKEPTPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKEPTTP 617

QY 661 KETAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKEPTTP 720

Db 618 KETAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKEPTTP 677

QY 721 PKPAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKEPTTP 780

Db 678 PKPAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKEPTTP 737

QY 781 TPKETAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKEPTTP 840

Db 738 TPKETAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKEPTTP 797

QY 841 AEPPTKALENSPKPGVPTTKPAATKPEMTTTAKDKTTERDLRTTTPETTTAAPKMTKET 900

Db 798 AEPPTKALENSPKPGVPTTKPAATKPEMTTTAKDKTTERDLRTTTPETTTAAPKMTKET 857

QY 901 ATTTEKTTESKIAATTTQVSTTTQDTPPKIITLKTTLAPKVTTKKLTITTEIMNKP 960

Db 858 ATTTEKTTESKIAATTTQVSTTTQDTPPKIITLKTTLAPKVTTKKLTITTEIMNKP 917

QY 961 BETAKPKDRATNSKATTPKPKOKPKAPKPKPTSTKPKTTPMPVRKPKTTPPKMTSTWPE 1020

Db 918 BETAKPKDRATNSKATTPKPKOKPKAPKPKPTSTKPKTTPMPVRKPKTTPPKMTSTWPE 977

QY 1021 LNPTSRIAEAMLOTTTRNCPQNSKLEVNPKSEDAAGBETPHMLLRPHVFMPEVTPD 1080

Db 978 LNPTSRIAEAMLOTTTRNCPQNSKLEVNPKSEDAAGBETPHMLLRPHVFMPEVTPD 1037

QY 1081 MDYLPRVNPQIILNPMLSDETNCNGKPVGLTTLRNGTLVAFRGHVMFLSPSPSPSP 1140

Db 1038 MDYLPRVNPQIILNPMLSDETNCNGKPVGLTTLRNGTLVAFRGHVMFLSPSPSPSPSP 1097

QY 1141 ARRTTEVWGIPIPDITVTRCNCEGKTFEFDKQSYWRFTNDIKDAGYKPKIFKFGGLTG 1200

Db 1098 ARRTTEVWGIPIPDITVTRCNCEGKTFEFDKQSYWRFTNDIKDAGYKPKIFKFGGLTG 1157

QY 1201 QIVAALSTAKYKNWPESVYFFKRGGSIOQYIKQEPVOKCPGRRPALNYPVYGMTQVRR 1260

Db 1158 QIVAALSTAKYKNWPESVYFFKRGGSIOQYIKQEPVOKCPGRRPALNYPVYGMTQVRR 1217

QY 1261 RRFERAIGSQHTTIRIQYSPARLAYQDKGVHLNEVKVSIILWRGLPNVVTSAISLPNIRK 1320

Db 1218 RRFERAIGSQHTTIRIQYSPARLAYQDKGVHLNEVKVSIILWRGLPNVVTSAISLPNIRK 1277

QY 1321 PDGYDYAFSKQYQYNNIDVPSRTARAITTRSGQTLSSKWYNCP 1363

Db 1278 PDGYDYAFSKQYQYNNIDVPSRTARAITTRSGQTLSSKWYNCP 1320

RESULT 8

US-07-757-022B-40  
; Sequence 40, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Geener, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cseri, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5951  
; INFORMATION FOR SEQ ID NO: 40:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1361 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-07-757-022B-40

Query Match 96.8%; Score 7044.6; DB 4; Length 1361;  
Best Local Similarity 94.0%; Pred. No. 5.1e-190;  
Matches 1320; Conservative 0; Mismatches 0; Indels 84; Gaps 2;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25  
Db 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 60  
QY 26 -----ELSCGRCFESPERGECDDAQQCKYDKCCPDYSCFAEAVHNPTSPSSKKAP 79  
Db 61 KKVCTAELSCGRCFESPERGECDDAQQCKYDKCCPDYSCFAEAVHNPTSPSSKKAP 120  
QY 80 PPSGASQTIKSTTKSPKPPNKKTKVIESEITEHSVSENQSSSSSSSSSSSTIW 139  
Db 121 PPSGASQTIKSTTKSPKPPNKKTKVIESEITE----- 156  
QY 140 KITSSKNSAANRELOKKLKVDKNKNTKKTPTKPPVVDVDEAGSLDNGDFKVTTPDTST 199

Db 157 -----VKDNKKNTKKKPTKPPVWDEAGSLDNGDFKVTTPDST 197  
QY 200 TQHNKYSTSKITTAKEINPRPSLPNSDTSKETSILVNVKETTVEYKETTITNNKQSTDG 259  
Db 198 TQHNKYSTSKITTAKEINPRPSLPNSDTSKETSILVNVKETTVEYKETTITNNKQSTDG 257  
QY 260 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTKEPTTTPKEPAS 319  
Db 258 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTKEPTTTPKEPAS 317  
QY 320 TTPKEPTTTPKSAPTTPKEPAPTTTTSKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 379  
Db 318 TTPKEPTTTPKSAPTTPKEPAPTTTTSKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 377  
QY 380 APPTTTSAPTTTKEPAPTTTTPKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 439  
Db 378 APPTTTSAPTTTKEPAPTTTTPKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 437  
QY 440 EPAPTAPKAPPTTTPKEPAPTTTTPKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 499  
Db 438 EPAPTAPKAPPTTTPKEPAPTTTTPKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 497  
QY 500 TTKSAPTTTKEPAPTTTTPKEPAPTTTTPKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 559  
Db 498 TTKSAPTTTKEPAPTTTTPKEPAPTTTTPKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 557  
QY 560 APAPTAPKEPAPTTTPKETAPTTTPKLTTPTEKLAAPTTPEKAPTTTPBELAPTTPEBPTT 619  
Db 558 APAPTAPKEPAPTTTPKETAPTTTPKLTTPTEKLAAPTTPEKAPTTTPBELAPTTPEBPTT 617  
QY 620 PEEPAPTTPKAAAPNTPKAPPTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTT 679  
Db 618 PEEPAPTTPKAAAPNTPKAPPTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTT 677  
QY 680 APPTPKAPKAPLAPTTTKEPTSTSDKAPPTTPKGTATTPKEPAPTTTPKEPAPTTTPKG 739  
Db 678 APPTPKAPKAPLAPTTTKEPTSTSDKAPPTTPKGTATTPKEPAPTTTPKEPAPTTTPKG 737  
QY 740 TAPPTLKEPAPTTTPKAPKAPLAPTTTKEPTSTSDKAPPTTPKGTATTPKEPAPTTTPKEP 799  
Db 738 TAPPTLKEPAPTTTPKAPKAPLAPTTTKEPTSTSDKAPPTTPKGTATTPKEPAPTTTPKEP 797  
QY 800 KPAPTTPPTTSEVSTPTTTPKPTTIHKSPDESTPELSAETPKALENSPKBGPVT 859  
Db 798 KPAPTTPPTTSEVSTPTTTPKPTTIHKSPDESTPELSAETPKALENSPKBGPVT 857  
QY 860 TKTPAATKPEMTTAKDKTTERDLRTTPETTTAAPKMTKETATTTTEKTTESKITATTIQV 919  
Db 858 TKTPAATKPEMTTAKDKTTERDLRTTPETTTAAPKMTKETATTTTEKTTESKITATTIQV 917  
QY 920 TSTTTQDTPPKITTLKTTLLAPKVTTKKLTITTEIMNKPEETAKPKDRATNSKATTPK 979  
Db 918 TSTTTQDTPPKITTLKTTLLAPKVTTKKLTITTEIMNKPEETAKPKDRATNSKATTPK 977  
QY 980 POKPTKAPKKPTS TKKPKTIPRVKPTTTPKMTSTMPNLNPTSRIAEMLOTTTRN 1039  
Db 978 POKPTKAPKKPTS TKKPKTIPRVKPTTTPKMTSTMPNLNPTSRIAEMLOTTTRN 1037  
QY 1040 QTPNSKLVEVNDKSDAGAGETHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS 1099  
Db 1038 QTPNSKLVEVNDKSDAGAGETHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS 1097  
QY 1100 DETNLCNGKPDVGLTTLRNGTLVAPRGHYFWMLSPPSPSPARRITEVWGIPSPIDTVFT 1159  
Db 1098 DETNLCNGKPDVGLTTLRNGTLVAPRGHYFWMLSPPSPSPARRITEVWGIPSPIDTVFT 1157  
QY 1160 RCNCEGKTTFFFKDSQYWRFTNDIKDAGYKPIFKGFGGLTGOIIVAALSTAKYKNWPESVY 1219  
Db 1158 RCNCEGKTTFFFKDSQYWRFTNDIKDAGYKPIFKGFGGLTGOIIVAALSTAKYKNWPESVY 1217  
QY 1220 FFKRGGSIQOYIYKQEPVQKCGRRPALNYPVYGMTQVRRRRRFFERAIGPSQTHIRIQY 1279

Db 1218 FFKRGGSIQOYIYKQEPVQKCGRRPALNYPVYGMTQVRRRRRFFERAIGPSQTHIRIQY 1277  
QY 1280 SPARLAYQDKVLHNEVKVSIILMRGLPNVVTSAISLENIRKPDGYDYIAFSKQOYINIDV 1339  
Db 1278 SPARLAYQDKVLHNEVKVSIILMRGLPNVVTSAISLENIRKPDGYDYIAFSKQOYINIDV 1337  
QY 1340 PSRTARAITTTRSGOTLSKWYNCP 1363  
Db 1338 PSRTARAITTTRSGOTLSKWYNCP 1361  
RESULT 9  
US-10-164-595-58  
; Sequence 58, Application US/10164595  
; Patent No. 6657054  
; GENERAL INFORMATION:  
; APPLICANT: Origene Technologies, Inc  
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides  
; FILE REFERENCE: IU 103 P1  
; CURRENT APPLICATION NUMBER: US/10/164,595  
; CURRENT FILING DATE: 2002-06-10  
; NUMBER OF SEQ ID NOS: 80  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 58  
; LENGTH: 1320  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-164-595-58  
Query Match 96.8%; Score 7043.7; DB 4; Length 1320;  
Best Local Similarity 96.6%; Pred. No. 5.1e-190;  
Matches 1317; Conservative 0; Mismatches 3; Indels 43; Gaps 1;  
QY 1 MAWKTLPIYLLLLSVFVIQQVSSQELSCGKCFESFERGREGCDCAQCKKYDKCCPDYE 60  
Db 1 MAWKTLPIYLLLLSVFVIQQVSSQELSCGKCFESFERGREGCDCAQCKKYDKCCPDYE 60  
QY 61 SFCAEVHNPSTSPSSKAPPPSGASQTIKSTTRGPKPPNKKTKKVIIESEITEHSVS 120  
Db 61 SFCAEVHNPSTSPSSKAPPPSGASQTIKSTTRGPKPPNKKTKKVIIESEITEHSVS 115  
QY 121 ENQESSSSSSSSSTIWKIKSSNKAANRELQKKLVKDNKKARTKKKPTPKPPVDE 180  
Db 116 -----VKDNKKARTKKKPTPKPPVDE 137  
QY 181 AGSGLDNGDFKVTTPDSTTQHNKYSTSKITTAKEINPRPSLPNSDTSKETSILVNVKE 240  
Db 138 AGSGLDNGDFKVTTPDSTTQHNKYSTSKITTAKEINPRPSLPNSDTSKETSILVNVKE 197  
QY 241 TTVEYKETTITNNKQSTDGKEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKG 300  
Db 198 TTVEYKETTITNNKQSTDGKEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKG 257  
QY 301 PALTTTKEPAPTTTTPKEPAPTTTTPKAPTTTKEPAPTTTTPKAPTTTTPKAPTTTTPK 360  
Db 258 PALTTTKEPAPTTTTPKEPAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPK 317  
QY 361 KEAPTTTTPKEPAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPK 420  
Db 318 KEAPTTTTPKEPAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPK 377  
QY 421 TTPKEPAPTTTTPKEPAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTT 480  
Db 378 TTPKEPAPTTTTPKEPAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTT 437  
QY 481 PAPTTTTSAPTTTTPKEPAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTT 540  
Db 438 PAPTTTTSAPTTTTPKEPAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTT 497  
QY 541 KEAPTTTTPKEPAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPK 600  
Db 498 KEAPTTTTPKEPAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPK 557

```
QY 601 PTTPELAPTTPEPTTTPPEPAPTTPKAAAPNTPKEPAPTTKPEAPTTKPEAPTTT 660
Db 558 PTTPELAPTTPEPTTTPPEPAPTTPKAAAPNTPKEPAPTTKPEAPTTKPEAPTTT 617
QY 661 KETAPTTKGTAPTTLKPEAPTTPKKAPKELAPTTTKEPTTSTSKPAPTTKGTAPTT 720
Db 618 KETAPTTKGTAPTTLKPEAPTTPKKAPKELAPTTTKEPTTSTCDKAPTTKGTAPTT 677
QY 721 PKEAPTTKPEAPTTPKGTAPTTLLKPEAPTTPKKAPKELAPTTTGTSTTSKAPTT 780
Db 678 PKEAPTTKPEAPTTPKGTAPTTLLKPEAPTTPKKAPKELAPTTTGTSTTSKAPTT 737
QY 781 TPKEAPTTKPEAPTTPKKAPTTTETPTTSEVSTTETTTKEPTTIHKSDDESPELS 840
Db 738 TPKEAPTTKPEAPTTPKKAPTTTETPTTSEVSTTETTTKEPTTIHKSDDESPELS 797
QY 841 ABPTPKALNSKPEKQVPTTKTAAATKPEMTTITAKDKTTTERDLRTTETTTAAPKMTKET 900
Db 798 ABPTPKALNSKPEKQVPTTKTAAATKPEMTTITAKDKTTTERDLRTTETTTAAPKMTKET 857
QY 901 ATTTEKTSKGTATTTQVSTTTQDTPPKITTLTKTTLAPKVTTKKTIITTEIMNKP 960
Db 858 ATTTEKTSKGTATTTQVSTTTQDTPPKITTLTKTTLAPKVTTKKTIITTEIMNKP 917
QY 961 BETAKPKDRATNSKATTPKPKPTKAPKPTSTKPKTWPVRVKPTTPTPRKMTSTMPD 1020
Db 918 BETAKPKDRATNSKATTPKPKPTKAPKPTSTKPKTWPVRVKPTTPTPRKMTSTMPD 977
QY 1021 LNPTSRIAEAMLTQTTTRPNQTPNSKLVEVNPKSEDAGGAGETPHMLLRPHVFMPEVTPD 1080
Db 978 LNPTSRIAEAMLTQTTTRPNQTPNSKLVEVNPKSEDAGGAGETPHMLLRPHVFMPEVTPD 1037
QY 1081 MDYLPRVNPQGIILNPMLSDETINICNGKPDVGLTTLRNGTLVAFRGHYFWMLSFFPSPSP 1140
Db 1038 MDYLPRVNPQGIILNPMLSDETINICNGKPDVGLTTLRNGTLVAFRGHYFWMLSFFPSPSP 1097
QY 1141 ARSITEVWGTPSPIDVTFECNCEGKTFEPKDSQYWRFTNDIKDAGVPKPIFKGFGGLTG 1200
Db 1098 ARSITEVWGTPSPIDVTFECNCEGKTFEPKDSQYWRFTNDIKDAGVPKPIFKGFGGLTG 1157
QY 1201 QIVAAALSTAKYKNWPESVVFKEKGGSIQQYIYKQEPVKQCGRRPALNYPVYGEMTQVRR 1260
Db 1158 QIVAAALSTAKYKNWPESVVFKEKGGSIQQYIYKQEPVKQCGRRPALNYPVYGEMTQVRR 1217
QY 1261 RRFERAIGPSQTHIRIQYSPARLAYQDKGVILHNEVKVSIILWRGLPNVVTSAISLPNIRK 1320
Db 1218 RRFERAIGPSQTHIRIQYSPARLAYQDKGVILHNEVKVSIILWRGLPNVVTSAISLPNIRK 1277
QY 1321 PDGYDYVAFSKDQYVNIIDVPSRTARAITTRSGQTLKSVWYNCP 1363
Db 1278 PDGYDYVAFSKDQYVNIIDVPSRTARAITTRSGQTLKSVWYNCP 1320
```

RESULT 10

```
US-07-757-022B-142
; Sequence 142, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserik, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 142:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1313 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-142
```

Query Match 96.2%; Score 7002; DB 4; Length 1313;

Best Local Similarity 96.3%; Pred. No. 7.5e-189;

Matches 1313; Conservative 0; Mismatches 0; Indels 50; Gaps 1;

```
QY 1 MAWKTPILYLLLSLVFVIQQVSSQELSKCKGRGCFESFERGREGDCAQCKKYDKCCPDYE 60
Db 1 MAWKTPILYLLLSLVFVIQQVSSQELSKCKGRGCFESFERGREGDCAQCKKYDKCCPDYE 60
QY 61 SFCAEVHNETSPSPSKKAPPPSGASQTIKSTTKRSPKPNKTKTKVIESEETEHSHVS 120
Db 61 SFCA-----EESHVS 70
QY 121 ENQESSSSSSSSSSSTTIWKIKSSKNSAANRELOKLVKDKNKNRTKKKPTPKPPVVDE 180
Db 71 ENQESSSSSSSSSSSTTIWKIKSSKNSAANRELOKLVKDKNKNRTKKKPTPKPPVVDE 130
QY 181 AGSGLONGDPKVTTPDTSTTQHNKVSTSPKITTAKPINRPPSLPPNSDTSKETSILTWNKE 240
Db 131 AGSGLONGDPKVTTPDTSTTQHNKVSTSPKITTAKPINRPPSLPPNSDTSKETSILTWNKE 190
QY 241 TTIVETKETTNNKQSTDGKEKTTSAKETQSIEKTSADLAPTSKVLAKTPKAEETTKG 300
Db 191 TTIVETKETTNNKQSTDGKEKTTSAKETQSIEKTSADLAPTSKVLAKTPKAEETTKG 250
QY 301 PALTTKPEPTTTPKPEASTTTPKEPTTTIKSAPTTTKEPAPTTTKSAPTTTKEPAPTTT 360
Db 251 PALTTKPEPTTTPKPEASTTTPKEPTTTIKSAPTTTKEPAPTTTKSAPTTTKEPAPTTT 310
QY 361 KEPAPTTTKEPAPTTTKEPAPTTTKSAPTTTKEPAPTTTPKKPAPTTTKEPAPTTTKEPTP 420
Db 311 KEPAPTTTKEPAPTTTKEPAPTTTKSAPTTTKEPAPTTTPKKPAPTTTKEPAPTTTKEPTP 370
QY 421 TTPKEPAPTTTKEPAPTTTKEPAPTTAPKKPAPTTTKEPAPTTTKEPAPTTTKEPSPTTKE 480
Db 371 TTPKEPAPTTTKEPAPTTTKEPAPTTAPKKPAPTTTKEPAPTTTKEPAPTTTKEPSPTTKE 430
QY 481 PAPTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 540
Db 431 PAPTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 490
```

```
QY 541 KEPAFTTPKBPATTTTKKPAPTAKEPAFTTPKETAFTTPKKLTPFTTPEKLAFTTPEKPA 600
Db 491 KEPAFTTPKBPATTTTKKPAPTAKEPAFTTPKETAFTTPKKLTPFTTPEKLAFTTPEKPA 550
QY 601 PTTPEELAPTTPEFTPTTPEPAFTTPKAAAPNTKPEPAFTTPKBPATTTPEKPAFTT 660
Db 551 PTTPEELAPTTPEFTPTTPEPAFTTPKAAAPNTKPEPAFTTPKBPATTTPEKPAFTT 610
QY 661 KETAPTTKGPATTLKEPAFTTPKAPKAPKELATTTKEPTSTTSDKPAFTT 720
Db 611 KETAPTTKGPATTLKEPAFTTPKAPKAPKELATTTKEPTSTTSDKPAFTT 670
QY 721 PKEPAFTTPKBPATTTKGPATTLKEPAFTTPKAPKAPKELATTTKGPSTTSDKPAFT 780
Db 671 PKEPAFTTPKBPATTTKGPATTLKEPAFTTPKAPKAPKELATTTKGPSTTSDKPAFT 730
QY 781 TPKEATTPKBPATTPKBPATTPETPTTSEVSTPTTKEPTTIHKSPDESTPELS 840
Db 731 TPKEATTPKBPATTPKBPATTPETPTTSEVSTPTTKEPTTIHKSPDESTPELS 790
QY 841 AEPTPKALENSKPEGVPTTKTAPATPEMTTAKDKTTERDITTTETTTAAPKMTKET 900
Db 791 AEPTPKALENSKPEGVPTTKTAPATPEMTTAKDKTTERDITTTETTTAAPKMTKET 850
QY 901 ATTTEKTTESKITATTTQVSTTTQDTTTPKKTTLKTTTLAPKVTTTKITTTTEIMNKP 960
Db 851 ATTTEKTTESKITATTTQVSTTTQDTTTPKKTTLKTTTLAPKVTTTKITTTTEIMNKP 910
QY 961 EETAKPKDRATNSKATTPKQKPTKAPKPTSTTKKPTMPRVRKPTTTPRKMWTSMPE 1020
Db 911 EETAKPKDRATNSKATTPKQKPTKAPKPTSTTKKPTMPRVRKPTTTPRKMWTSMPE 970
QY 1021 LNFTSRIAEAMLQTTTRPNQTPNSKLVENPKSEDAGAGETPHMLLRHVFMPVEVTPD 1080
Db 971 LNFTSRIAEAMLQTTTRPNQTPNSKLVENPKSEDAGAGETPHMLLRHVFMPVEVTPD 1030
QY 1081 MDYLPVPVNGIILINPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPEPSPSP 1140
Db 1031 MDYLPVPVNGIILINPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPEPSPSP 1090
QY 1141 ARRIEIVGWIPSIDVTRCNCCEGKTFPKDSQYWRFTNDIKDAGYKPEIFKGFGLTG 1200
Db 1091 ARRIEIVGWIPSIDVTRCNCCEGKTFPKDSQYWRFTNDIKDAGYKPEIFKGFGLTG 1150
QY 1201 QIVAAALSTAKYKNWPESVVFVFFKGGSTQQVIYKQEPVQKCPGRPALNYPVYGEMTQVRR 1260
Db 1151 QIVAAALSTAKYKNWPESVVFVFFKGGSTQQVIYKQEPVQKCPGRPALNYPVYGEMTQVRR 1210
QY 1261 RRFERAIGPSQTHIRIQYSPARLAYQDKGVHLHNEVKVSIILWRGLPNVVTSAISLBNIRK 1320
Db 1211 RRFERAIGPSQTHIRIQYSPARLAYQDKGVHLHNEVKVSIILWRGLPNVVTSAISLBNIRK 1270
QY 1321 PDGYDYAFSKQYNYNDVPSRTARAITTRSGOTLSKWNVNC 1363
Db 1271 PDGYDYAFSKQYNYNDVPSRTARAITTRSGOTLSKWNVNC 1313
```

RESULT 11

```
US-07-757-022B-48
; Sequence 48, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
```

```
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1354 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-48
```

```
Query Match 96.0%; Score 6987.9; DB 4; Length 1354;
Best Local Similarity 93.5%; Pred. No. 2e-188;
Matches 1313; Conservative 0; Mismatches 0; Indels 91; Gaps 2;

QY 1 MAWKTLPIYLLLSLVFVIQVSSQ----- 25
Db 1 MAWKTLPIYLLLSLVFVIQVSSQDLSSCAGRCGEGYSDATCNCYNCQHYMECCPDF 60
QY 26 -----ELSCGRCFESFERGECDCDQCKKYDKCCPDYESFCAEVHNPTSPSSKKAP 79
Db 61 KRVCVTRAEISCKGRCFESFERGECDCDQCKKYDKCCPDYESFCA----- 105
QY 80 PPSGASQTIKSTTKRSPKPPNKKTKKVIETEEHSVSENQBSSESSSSSSSSSSSTIW 139
Db 106 -----EEHSVSENQBSSESSSSSSSSSSSTIW 130
QY 140 KIKSKNSAANRELQKLVKONKNRTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDTST 199
Db 131 KIKSKNSAANRELQKLVKONKNRTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDTST 190
QY 200 TOHNKVSTSPKLTITAKPINPREPLPNSDTSKTSITVKNKETTIVETKETTITNKQSTDG 259
Db 191 TOHNKVSTSPKLTITAKPINPREPLPNSDTSKTSITVKNKETTIVETKETTITNKQSTDG 250
QY 260 KEKTTSAKETOSIEKTSARDLAPTSKVLAKPTPKAETTTKGPAITTPKEPTTTPKEPAS 319
Db 251 KEKTTSAKETOSIEKTSARDLAPTSKVLAKPTPKAETTTKGPAITTPKEPTTTPKEPAS 310
QY 320 TTPKEPTTTPKSAPTTPKBPATTTKSAPTTPKBPATTTTPKEPTTTPKEPTTTPKEPT 379
Db 311 TTPKEPTTTPKSAPTTPKBPATTTKSAPTTPKBPATTTTPKEPTTTPKEPTTTPKEPT 370
QY 380 APTTTKSAPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPAT 439
Db 371 APTTTKSAPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPAT 430
```









QY 241 TTVETKEITTTNKQSTGDKETTSKAKTOSIEKTSKDLAPTSKVLAKPTPKAETTTKG 300  
Db 148 TTVETKEITTTNKQSTGDKETTSKAKTOSIEKTSKDLAPTSKVLAKPTPKAETTTKG 207  
QY 301 PALTTTKEPTPTTPKEPASTTPKEPTPTTIKSAPTTTPKEPAPTTPKSAPTTTPKEPAPTTP 360  
Db 208 PALTTTKEPTPTTPKEPASTTPKEPTPTTIKSAPTTTPKEPAPTTPKSAPTTTPKEPAPTTP 267  
QY 361 KEPAPTTPKEPAPTTPKEPAPTTPKGAPTTPKEPAPTTPKKAAPTTPKEPAPTTPKEPAPTTP 420  
Db 268 KEPAPTTPKEPAPTTPKEPAPTTPKGAPTTPKEPAPTTPKKAAPTTPKEPAPTTPKEPAPTTP 327  
QY 421 TTPKEPAPTTPKEPAPTTPKEPAPTTPKGAPTTPKEPAPTTPKKAAPTTPKEPAPTTPKEPAPTTP 480  
Db 328 TTPKEPAPTTPKEPAPTTPKEPAPTTPKGAPTTPKEPAPTTPKKAAPTTPKEPAPTTPKEPAPTTP 387  
QY 481 PAPTTTKEPAPTTPKEPAPTTPKGAPTTPKEPAPTTPKKAAPTTPKEPAPTTPKEPAPTTP 540  
Db 388 PAPTTTKEPAPTTPKEPAPTTPKGAPTTPKEPAPTTPKKAAPTTPKEPAPTTPKEPAPTTP 447  
QY 541 KEPAPTTPKEPAPTTPKEPAPTTPKGAPTTPKEPAPTTPKKAAPTTPKEPAPTTPKEPAPTTP 600  
Db 448 KEPAPTTPKEPAPTTPKEPAPTTPKGAPTTPKEPAPTTPKKAAPTTPKEPAPTTPKEPAPTTP 507  
QY 601 PTTPEELAPTTPEPTPTTPKEPAPTTPKGAPTTPKEPAPTTPKKAAPTTPKEPAPTTPKEPAPTTP 660  
Db 508 PTTPEELAPTTPEPTPTTPKEPAPTTPKGAPTTPKEPAPTTPKKAAPTTPKEPAPTTPKEPAPTTP 567  
QY 661 KETAPTTPKEPAPTTPKEPAPTTPKGAPTTPKEPAPTTPKKAAPTTPKEPAPTTPKEPAPTTP 720  
Db 568 KETAPTTPKEPAPTTPKEPAPTTPKGAPTTPKEPAPTTPKKAAPTTPKEPAPTTPKEPAPTTP 627  
QY 721 PKEPAPTTPKEPAPTTPKEPAPTTPKGAPTTPKEPAPTTPKKAAPTTPKEPAPTTPKEPAPTTP 780  
Db 628 PKEPAPTTPKEPAPTTPKEPAPTTPKGAPTTPKEPAPTTPKKAAPTTPKEPAPTTPKEPAPTTP 687  
QY 781 TPKETAPTTPKEPAPTTPKEPAPTTPKGAPTTPKEPAPTTPKKAAPTTPKEPAPTTPKEPAPTTP 840  
Db 688 TPKETAPTTPKEPAPTTPKEPAPTTPKGAPTTPKEPAPTTPKKAAPTTPKEPAPTTPKEPAPTTP 747  
QY 841 AETPKALENSKEPAGVPTTKPAATKPEMTTTAKDKITTERDLATTPETTTAAPKMTKET 900  
Db 748 AETPKALENSKEPAGVPTTKPAATKPEMTTTAKDKITTERDLATTPETTTAAPKMTKET 807  
QY 901 ATTEKTESKITATTTQVSTTTQDTPPKITLTKTLTAPKVTITTKITITTEIMNKP 960  
Db 808 ATTEKTESKITATTTQVSTTTQDTPPKITLTKTLTAPKVTITTKITITTEIMNKP 867  
QY 961 EETAKPKDRATNSKATTPKOKPTKAPKPTSTKPKTTPRVRKPTTPRKMSTWPE 1020  
Db 868 EETAKPKDRATNSKATTPKOKPTKAPKPTSTKPKTTPRVRKPTTPRKMSTWPE 927  
QY 1021 LNPTSRIAEAMLTQTRPNQTPNSKLVENPKSBDAGAGETPHMLLRPHVFMVEVTPD 1080  
Db 928 LNPTSRIAEAMLTQTRPNQTPNSKLVENPKSBDAGAGETPHMLLRPHVFMVEVTPD 987  
QY 1081 MDYLPVFNQIIINPMLSDETNINCGRPVDGLTTLNGLTAVARGHVFWMVSPSPSP 1140  
Db 988 MDYLPVFNQIIINPMLSDETNINCGRPVDGLTTLNGLTAVARGHVFWMVSPSPSP 1047  
QY 1141 ARRITEVWGIPSPIDTVPTRCNCEGKTPFFKDSQVWRFTNDIKDAGYKPIFKFGGLTG 1200  
Db 1048 ARRITEVWGIPSPIDTVPTRCNCEGKTPFFKDSQVWRFTNDIKDAGYKPIFKFGGLTG 1107  
QY 1201 QIVAALSTAKYKNWPESVYFFKRGSGIOQYIKOEPVOKCPGRPALNYPYVGMTQVRR 1260  
Db 1108 QIVAALSTAKYKNWPESVYFFKRGSGIOQYIKOEPVOKCPGRPALNYPYVGMTQVRR 1167  
QY 1261 RFRERAGPSOTHIRIOYSPARLAYQDKGVHNEVKVSIILWRGLPNVVTSAISLPNIRK 1320  
Db 1168 RFRERAGPSOTHIRIOYSPARLAYQDKGVHNEVKVSIILWRGLPNVVTSAISLPNIRK 1227  
QY 1321 PDGYDYAFSKQYYNIDVPSRTARAITTRSGQTLISKVYNCP 1363

Db 1228 PDGYDYAFSKQYYNIDVPSRTARAITTRSGQTLISKVYNCP 1270  
RESULT 14  
US-07-757-022B-42  
; Sequence 42, Application US/07757022B  
; Patent No. 6431142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM: disk  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cseri, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 42:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1311 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-07-757-022B-42  
Query Match 93.2%; Score 6780.6; DB 4; Length 1311;  
Best Local Similarity 90.5%; Pred. No. 1.2e-182;  
Matches 1270; Conservative 0; Mismatches 0; Indels 134; Gaps 2;  
QY 1 MAWKTLPIYLLLSVFIQQVSSQ----- 25  
Db 1 MAWKTLPIYLLLSVFIQQVSSQDLSSCAGRCGYSRDATCNCYDNCQHYMECCPDF 60  
QY 26 -----ELSCKGRCFESFERGECDCDAQCKKYDKCCPDYEFSCAEVHNPTSPSSKKAP 79  
Db 61 KRVCCTAELSCKGRCFESFERGECDCDAQCKKYDKCCPDYEFSCAE----- 106  
QY 80 PPGASQTIKSTTKRSPKPPNKKTKKVIIESEEITEHSVSENQESSSSSSSSSSSTIW 139  
Db 107 ----- 106

QY 140 KIKSSKNSAANRELQKKLVKDNKNRRTKKKPTPKPPVVDGAGSLONGDFKVTTPDST 199  
Db 107 -----VKDNKNRRTKKKPTPKPPVVDGAGSLONGDFKVTTPDST 147  
QY 200 TOHNKUSTSPKIIITAKPINRPSLPNSDTSKETSLSLVNKETTIVETKETTNNKQSTG 259  
Db 148 TOHNKUSTSPKIIITAKPINRPSLPNSDTSKETSLSLVNKETTIVETKETTNNKQSTG 207  
QY 260 KEKTTSAKETQSIIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTKEPTTTPKBPAS 319  
Db 208 KEKTTSAKETQSIIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTKEPTTTPKBPAS 267  
QY 320 TTPKEPTPTTIKSAPTTTPKEPAPTTTTSAPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEP 379  
Db 268 TTPKEPTPTTIKSAPTTTPKEPAPTTTTSAPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEP 327  
QY 380 APITTTKSAPTTTPKEPAPTTTPKAPATTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 439  
Db 328 APITTTKSAPTTTPKEPAPTTTPKAPATTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 387  
QY 440 EPAPTAPKAPATTPKEPAPTTTPKEPAPTTTTPKAPATTPKEPAPTTTTPKAPATTPKEP 499  
Db 388 EPAPTAPKAPATTPKEPAPTTTPKEPAPTTTTPKAPATTPKEPAPTTTTPKAPATTPKEP 447  
QY 500 TTKSAPTTTPKEPAPTTTTPKEPAPTTTPKAPATTPKEPAPTTTPKEPAPTTTTPKAP 559  
Db 448 TTKSAPTTTPKEPAPTTTTPKEPAPTTTPKAPATTPKEPAPTTTPKEPAPTTTTPKAP 507  
QY 560 APAPTAPKEPAPTTTPKEPAPTTTPKAPATTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 619  
Db 508 APAPTAPKEPAPTTTPKEPAPTTTPKAPATTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 567  
QY 620 PERPAPTTPKAAAPNTPKAPATTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 679  
Db 568 PERPAPTTPKAAAPNTPKAPATTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 627  
QY 680 APPTPKAPKELAPTTTKEPTSTTSKAPATTPKGTAPTTTPKEPAPTTTPKEPAPTTTPK 739  
Db 628 APPTPKAPKELAPTTTKEPTSTTSKAPATTPKGTAPTTTPKEPAPTTTPKEPAPTTTPK 687  
QY 740 TAPTTTKEPAPTTTPKAPKELAPTTTKEPTSTTSKAPATTPKGTAPTTTPKEPAPTTTPK 799  
Db 688 TAPTTTKEPAPTTTPKAPKELAPTTTKEPTSTTSKAPATTPKGTAPTTTPKEPAPTTTPK 747  
QY 800 KPAPTTPETPTTSEVSTPTTKEPTTHKSDESTPELSAETTPKALENSKPECVPT 859  
Db 748 KPAPTTPETPTTSEVSTPTTKEPTTHKSDESTPELSAETTPKALENSKPECVPT 807  
QY 860 TKTPAATKPEMTTAKDKTTERDLRTTPTTAAAPKMTKETATTTTEKTTESKITATTTOV 919  
Db 808 TKTPAATKPEMTTAKDKTTERDLRTTPTTAAAPKMTKETATTTTEKTTESKITATTTOV 867  
QY 920 TSTTTQDTTTPFKITLKITTLAPKVTITTKITITTIMNKPEBETAKPKORAINSKATTPK 979  
Db 868 TSTTTQDTTTPFKITLKITTLAPKVTITTKITITTIMNKPEBETAKPKORAINSKATTPK 927  
QY 980 POKPTKAPKKPTSTKPKMTWPRVRKPTTTPKMTSTWPELNPSTWPELNPSTWPELNPSTW 1039  
Db 928 POKPTKAPKKPTSTKPKMTWPRVRKPTTTPKMTSTWPELNPSTWPELNPSTWPELNPSTW 987  
QY 1040 QTPNSKLVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMDYLPRVNOGIIINPMLS 1099  
Db 988 QTPNSKLVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMDYLPRVNOGIIINPMLS 1047  
QY 1100 DETNICNGPVDGLTLRLNGTLVAFRGHYFVWMLSPSPSPSPARRITEVWGISPIDTVP 1159  
Db 1048 DETNICNGPVDGLTLRLNGTLVAFRGHYFVWMLSPSPSPSPARRITEVWGISPIDTVP 1107  
QY 1160 RCNCEGKTFEKKDSQVWRFTNDIKAGYKPKIFKFGGLTGQIVAAALSTAKYKNWPESV 1219  
Db 1108 RCNCEGKTFEKKDSQVWRFTNDIKAGYKPKIFKFGGLTGQIVAAALSTAKYKNWPESV 1167

QY 1220 FFKGGSSIQOYIYKQEPVQKCPGRPALNVPVYGMTQVRRRERAIQPSQTHIRIQY 1279  
Db 1168 FFKGGSSIQOYIYKQEPVQKCPGRPALNVPVYGMTQVRRRERAIQPSQTHIRIQY 1227  
QY 1280 SPARLAYQDKGVLHNEVKVSLWGLPNVVTSAISLPNIRKPDGVDYVYAFSKDOYVNDV 1339  
Db 1228 SPARLAYQDKGVLHNEVKVSLWGLPNVVTSAISLPNIRKPDGVDYVYAFSKDOYVNDV 1287  
QY 1340 PSRTARAITTRSGQTLSSKQWYNCP 1363  
Db 1288 PSRTARAITTRSGQTLSSKQWYNCP 1311  
RESULT 15  
US-07-757-022B-104  
; Sequence 104, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 104:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1140 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-07-757-022B-104  
Query Match 80.0%; Score 5820.9; DB 4; Length 1140;  
Best Local Similarity 96.4%; Pred. No. 9.2e-156;  
Matches 1099; Conservative 0; Mismatches 0; Indels 41; Gaps 1;  
QY 1 MAWKTLPTLYLLLLLSVFVQQVSSQ-----25

Search completed: October 13, 2004, 11:59:15  
Job time : 29.693 secs

Db	1	MAWKTLPIYLLLLLSVFIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHMECCPDF	60
QY	26	-----ELSCGRCRCPESPERGECDCDAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP	79
Db	61	KRVCTAELSCGRCRCPESPERGECDCDAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP	120
QY	80	PPSGASQTIKSTTKRSPKPNKKTKVIESEITEHSVSENQESSSSSSSSSSSTIW	139
Db	121	PPSGASQTIKSTTKRSPKPNKKTKVIESEITEHSVSENQESSSSSSSSSSSSSTIW	180
QY	140	KIKSSKNSAANRELQKKLVKDKNKNRTKKPTPKPPVVDEAGSGLDNGDPKVITPTST	199
Db	181	KIKSSKNSAANRELQKKLVKDKNKNRTKKPTPKPPVVDEAGSGLDNGDFKVITPTST	240
QY	200	TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVNKETTVEKITTNNKQTSNDG	259
Db	241	TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVNKETTVEKITTNNKQTSNDG	300
QY	260	KEKTTSAKETQSIKTSKADLAETSKVLAKPTPKAEITTKGPALTTPKETPTTKPEPAS	319
Db	301	KEKTTSAKETQSIKTSKADLAETSKVLAKPTPKAEITTKGPALTTPKETPTTKPEPAS	360
QY	320	TTPEKPTPTTIKSAPITPKBPAPITTKSAPITPKBPAPITTKBPAPITTKBPAPITTKBP	379
Db	361	TTPEKPTPTTIKSAPITPKBPAPITTKSAPITPKBPAPITTKBPAPITTKBPAPITTKBP	420
QY	380	APTNTKSAPITPKBPAPITTKBPAPITTKBPAPITTKBPAPITTKBPAPITTKBPAPIT	439
Db	421	APTNTKSAPITPKBPAPITTKBPAPITTKBPAPITTKBPAPITTKBPAPITTKBPAPIT	480
QY	440	EPAPTAPKKAPITPKBPAPITTKBPAPITTKBPAPITTKBPAPITTKBPAPITTKBPAPIT	499
Db	481	EPAPTAPKKAPITPKBPAPITTKBPAPITTKBPAPITTKBPAPITTKBPAPITTKBPAPIT	540
QY	500	TTKSAPITPKBPAPITTKBPAPITTKBPAPITTKBPAPITTKBPAPITTKBPAPITTKBP	559
Db	541	TTKSAPITPKBPAPITTKBPAPITTKBPAPITTKBPAPITTKBPAPITTKBPAPITTKBP	600
QY	560	APTAPKBPAPITPKETAPTPKKLTPTTPEKLAETTPPEKLAETTPPEKLAETTPPEKLAET	619
Db	601	APTAPKBPAPITPKETAPTPKKLTPTTPEKLAETTPPEKLAETTPPEKLAETTPPEKLAET	660
QY	620	PEBPAPITPKAAAPNTPKBPAPITTKBPAPITTKBPAPITTKBPAPITTKBPAPITTKBPAPIT	679
Db	661	PEBPAPITPKAAAPNTPKBPAPITTKBPAPITTKBPAPITTKBPAPITTKBPAPITTKBPAPIT	720
QY	680	APTTPKKPAPKELAPITTKETSTTSKBPAPITTKGTAETTPKBPAPITTKBPAPITTKBPAPIT	739
Db	721	APTTPKKPAPKELAPITTKETSTTSKBPAPITTKGTAETTPKBPAPITTKBPAPITTKBPAPIT	780
QY	740	TAPTTLKBPAPITPKKAPKELAPITTKGPTSTTSKBPAPITTKGTAETTPKBPAPITTKBPAPIT	799
Db	781	TAPTTLKBPAPITPKKAPKELAPITTKGPTSTTSKBPAPITTKGTAETTPKBPAPITTKBPAPIT	840
QY	800	KPAPTTPPTTSEVSTPTTKETPTTIHKSPDESTPELSAETPKALENSPKBPAPITPK	859
Db	841	KPAPTTPPTTSEVSTPTTKETPTTIHKSPDESTPELSAETPKALENSPKBPAPITPK	900
QY	860	TKTPAANKPEMTTAKOKTTERDLRTPETTTAAAPKMTKETATTTETKITESKITATTQV	919
Db	901	TKTPAANKPEMTTAKOKTTERDLRTPETTTAAAPKMTKETATTTETKITESKITATTQV	960
QY	920	TSTTTQDTPPKITTLKTTLLAPKVTITTKIITTEIMNKPEETAKPKDRATNSKATTPK	979
Db	961	TSTTTQDTPPKITTLKTTLLAPKVTITTKIITTEIMNKPEETAKPKDRATNSKATTPK	1020
QY	980	POKPTKAPKPTSTKKPKTMEVRKPTTTPRKNSTMPNELNPTSRIAEAMLOTTTRPN	1039
Db	1021	POKPTKAPKPTSTKKPKTMEVRKPTTTPRKNSTMPNELNPTSRIAEAMLOTTTRPN	1080
QY	1040	QTPNSKLVENPKSEDAGAEGETPHMLLRPHVFMPEVTPDMXYLPRVFNQGIINPMLS	1099
Db	1081	QTPNSKLVENPKSEDAGAEGETPHMLLRPHVFMPEVTPDMXYLPRVFNQGIINPMLS	1140